AWARD NUMBER: W81XWH-17-1-0093

TITLE: Mutations in the Spliceosomal Gene ZRSR2 in Myelodysplastic Syndromes

PRINCIPAL INVESTIGATOR: H. Phillip Koeffler

CONTRACTING ORGANIZATION: Cedars-Sinai Medical Center
Los Angeles, CA 90048

REPORT DATE: April 2019

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
April 2019	Annual	1 APR 2018 – 31 MAR 2019

4. TITLE AND SUBTITLE

Mutations in the Spliceosomal Gene ZRSR2 in Myelodysplastic Syndromes

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-17-1-0093

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

H. Phillip Koeffler

E-Mail: h.koeffler@cshs.org

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Cedars-Sinai Medical Center
8700 Beverly Blvd.
Los Angeles, CA 90048-1804

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Myelodysplastic syndromes (MDS) are characterized by recurrent somatic alterations often affecting components of RNA splicing machinery. Mutations of splice factors SF3B1, SRSF2, ZRSR2 and U2AF1 occur in >50% of MDS. To assess the impact of spliceosome mutations on splicing and to identify common pathways/genes affected by distinct mutations, we performed RNA-sequencing of 32 MDS bone marrow samples harboring solely spliceosome mutations (including SF3B1, SRSF2 P95 or small deletions, U2AF1 S34 or Q157 hotspot mutations and ZRSR2 truncating alterations), but devoid of other co-occurring mutations. We uncover the landscape of splicing alterations in each splice factor mutant MDS and demonstrate that SRSF2 deletions cause highest number of splicing alterations compared with other spliceosome mutations. Although the mis-spliced events observed in different splice factor mutations were largely non-overlapping, a subset of genes, including *EZH2*, were aberrantly spliced in multiple mutant groups. Pathway analysis revealed dysregulated biological processes including RNA splicing and transport as well as several signaling cascades, which were observed repeatedly in different mutant groups, suggesting converging biological consequences downstream of distinct spliceosome mutations.

15. SUBJECT TERMS

ZRSR2, myelodysplastic syndromes (MDS), splicing factor mutations, U12-type intron splicing

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
Unclassified	Unclassified	Unclassified	Unclassified	63	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

Page

1. INTRODUCTION

Myelodysplastic syndromes (MDS) are a group of clonal hematopoietic disorders characterized by ineffective hematopoiesis resulting in peripheral blood cytopenias of myeloid lineages, including anemia and neutropenia, and a tendency to evolve into acute myeloid leukemia (AML). Limited understanding of the pathology of MDS has been a major impediment for development of novel therapies. The recent discovery that splicing factor genes are the most commonly mutated genes in MDS revealed that dysfunction of RNA splicing is a crucial pathogenic pathway in this disease. Most of these mutations occur in *SF3B1*, *U2AF1*, *SRSF2* and *ZRSR2*. Although several studies begun to elucidate the effects of each of these mutations on splicing and hematopoiesis, many questions remain open. In contrast to mutations of *SF3B1*, *SRSF2* and *U2AF1* (heterozygous missense mutations at restricted residues, conferring an alteration of function), mutations of the X-linked *ZRSR2* gene occur throughout the transcript and usually occur as nonsense or frameshift mutations (pattern consistent with loss-of-function). Additionally, ZRSR2 is predominantly important for U12 intron splicing, while the other three factors regulate major intron splicing. Despite these differences, mutations in *ZRSR2* are largely mutually exclusive with mutations of *SF3B1*, *SRSF2* and *U2AF1*; and moreover, like the other splicing mutations, leukemic cells with mutated *ZRSR2* appear to be preferentially sensitive to pharmacological inhibition of splicing. Thus, elucidating ZRSR2 function in MDS will provide important insights into the role of aberrant splicing in the pathogenesis of MDS, and will help development of specific therapies targeting cells bearing splicing mutations.

KEYWORDS

ZRSR2, myelodysplastic syndromes (MDS), splicing factor mutations, U12-type intron splicing

3. ACCOMPLISHMENTS

- **Major goals of the project**

Specific Aim 1: Determine how ZRSR2 mutations dysregulate downstream pathways, and contribute to molecular features of myelodysplastic cells.

Specific Aim 2: Study biological implications of loss of ZRSR2 in normal and MDS marrow cells.

Specific Aim 3: Investigate the role of ZRSR2 in myeloid transformation using *Zrsr2/Tet2*-deficent murine model.

- **Accomplishments under these goals**

Specific Aim 1: Determine how ZRSR2 mutations dysregulate downstream pathways and contribute to molecular features of myelodysplastic cells.

1A. Activation of DNA damage response. RNA-seq data show that multiple splicing factor mutations, including *ZRSR2*, widely target genes involved in DNA damage response, making DNA damage response a top candidate for a convergent mechanism in MDS with spliceosome mutations. To examine whether *ZRSR2*-deficient cells exhibit increased levels of DNA damage, we measured DNA damage with a pan-DNA damage marker H2AX-pSer139 (γH2AX) antibody, in leukemic cell lines (TF1, K562) stably transfected with either ZRSR2-shRNA or Cont-shRNA. Immunostaining showed a small increase in γH2AX signals in *ZRSR2*-knockdown cells compared to control cells (Fig.1).

1B. Analysis of R-loops levels. Elevated levels of DNA damage in cells with certain splicing factor mutations is mediated at least in part through induction of R-loops in a splicing independent manner. To determine whether increased R-loop formation occurs in *ZRSR2*-knockdown cells, we measured R-loops with S9.6, an antibody that specifically recognizes RNA/DNA hybrids. Genomic DNA from stably transfected ZRSR2-shRNA or Cont-shRNA cells was immunoprecipitated with monoclonal S9.6 antibody, spotted on nitrocellulose membranes and stained with a secondary antibody. Results showed higher levels of R-loops in HEK293T, K562 and TF1 *ZRSR2*-

knockdown cells (Fig. 2). To strengthen these results, we used HEK293T cells stably transfected with either wild-type (WT) or mutant (D210N or WKKD) RNASEH1, an enzyme that specifically degrades the RNA in R-loops. D210N mutation abolishes catalytic activity of RNASEH1; WKKD mutation abolishes both binding and catalytic activities of RNASEH1 and is used for negative control. We generated HEK293T cells stably expressing either V5-tagged RNASEH1 WT, D210N, or WKKD (vectors were a gift from Dr. D Zhang) and further transduced the cells with either ZRSR2-shRNA or Cont-shRNA (Fig. 3). Dot blot analysis with S9.6 showed elevated R-loops in ZRSR2-depleted RNASEH1 WT cells, which further increased in D210N expression cells, and not in WKKD expression cells (Fig. 4). We are now using the ZRSR2-depleted/RNASEH1 expressing cells to map the location of R-loops induced by loss of ZRSR2 with R-loop chromatin immunoprecipitation (R-ChIP), a technique in which R-loops are immunoprecipitated by a RNASEH1(D210N), followed by standard ChIP-seq.

1C. ATR activation. ATR-CHK1 and ATM-CHK2 are two crucial pathways controlling the DNA damage response pathway. Induction of R-loops by mutated splicing factor activates the ATR but not ATM pathway. Cells with *U2AF1* mutation show increased sensitivity to ATR inhibition, which is depended on the presence of R loops. Similarly, we found that ATR inhibitor VE-821 reduced the viability of *ZRSR2*-knockdown TF1 cells more than that of control cells (Fig. 5). We are now confirming these results in other *ZRSR2*-knockdown human cell lines. In addition, we are using ATR inhibitors in combination with MDS-relevant drugs (e.g. spliceosome modulators and hypomethylating agents) to test for synergy.

1D. Activation of innate immune signaling. Activation of innate immune signaling was proposed as another unifying mechanism mediating the effects of splicing factor mutations in MDS. To test whether loss of ZRSR2 renders cells more sensitive to inflammatory stimuli and NF-κB activation, we treated TF1 *ZRSR2* WT and knockdown (shRNA) cells with either lipopolysaccharide (LPS) or tumor necrosis factor α (TNFα) and measured levels of p65 phosphorylation (p-p65), a marker for NF-kB activation. Western blot analysis showed increased p-p65 levels in *ZRSR2* knockdown cells compared with control cells following TNFα but not LPS treatment (Fig. 6). The reason for specific induction by TNFα and not LPS is unclear. A possible explanation is that loss of ZRSR2 induces changes in genes that respond to TNFα but not to LPS stimulation.

1E. Distinct and convergent consequences of different splicing factor mutations. Though several studies identified common mechanisms in MDS patients with different splicing factor mutations, these studies either understudied or completely excluded *ZRSR2*. Also, many of the patient samples used contained additional mutations in known myeloid neoplasms-driver genes, making it difficult to link spliceosome mutations with downstream pathways. This prompted us to undertake a side-by-side analysis of carefully curated MDS samples. We performed RNA-Seq on a cohort of 24 MDS bone marrow samples harboring solely spliceosome mutations (*SF3B1*, *SRSF2* P95 or frameshift deletion, *U2AF1* S34 or Q157 hotspot mutations, and *ZRSR2* truncating alterations) and devoid of other frequently co-occurring mutations, and 8 samples without spliceosome mutation as control, using the Illumina TruSeq Sample preparation kit. Results from this study are presented in a manuscript we are currently writing (Appendix 1).

Specific Aim 2: Study biological implications of loss of ZRSR2 in normal and MDS marrow cells.

2A. Zrsr2 knockout mice. To investigate the role of ZRSR2 in normal hematopoiesis and leukemogenesis, we generated a *Zrsr2* knockout (KO) murine model. Male *Zrsr2* KO mice were obtained at expected frequency and did not display any apparent morphological abnormalities. RNA-seq analysis of total marrow revealed moderate increased mis-splicing of U12-type introns in KO cells compared with WT cells. Detail phenotypic analysis of hematopoietic compartments of both young (8-weeks old) and old (1-year old) *Zrsr2* KO mice showed no significant difference compared to their WT counterparts, suggesting that the increased mis-splicing is insufficient to drive dysplasia. For example, no difference in peripheral blood cell counts, proportion and absolute numbers of hematopoietic stem cells (HSCs) including long-term and short-term HSCs cells, or the numbers of myeloid precursors was noted (Fig. 7). Similarly, frequency of mature myeloid cells in bone marrow and spleen of *Zrsr2* KO mice were largely conserved (data not shown). Competitive bone marrow transplantation assays with *Zrsr2* KO hematopoietic progenitor cells showed that their multi-lineage reconstitution potential is comparable with that of WT cells (Fig. 8). Despite the lack of an obvious disease phenotype, our preliminary studies indicate that

analysis of mis-splicing events in bone marrow of *Zrsr2* KO mice can provide valuable insights into the importance of ZRSR2-induced splicing in hematopoietic differentiation. Particularly, when combined with data from human cells, the murine data can point to common downstream pathways and help prioritize genes for further studies.

2B. Gene expression and splicing alterations in murine *Zrsr2* knockout myeloid progenitors. Our preliminary splicing analysis of total bone morrow suggested that analyzing splicing alterations in *Zrsr2* KO murine myeloid precursors can be used to identify cell-type specific ZRSR2 downstream targets and pathways. To investigate the consequences of *Zrsr2*-deficiency on splicing in specific hematopoietic lineages, we will perform RNA-seq on FACS-sorted hematopoietic populations from bone marrow of *Zrsr2* KO mice. In preliminary analysis on sorted common myeloid progenitors (CMPs), granulocyte monocyte progenitor (GMPs), and megakaryocyte erythroid progenitor (MEPs), we observed increased retention of U12-type introns in *Zrsr2* KO cells compared with WT cells (Fig. 9). In contrast retention of U12-type introns in murine embryonic fibroblasts (MEFs) reminded largely unchanged. there

2C. ZRSR2 interacting partners. To identify ZRSR2 interacting proteins, we fused the C-terminus and N-terminus of ZRSR2 to both biotin ligase (BirA*) and FLAG (V9595 and V9596, respectively). We transfected ZRSR2-BirA*-FLAG or BirA*-FLAG into Flp-In T-REx 293 cells and established stable and inducible isogenic cell lines ±ZRSR2-BirA*-FLAG. Inducible expression of C-terminus (V9595) and N-terminus (V9596) tagged-ZRSR2 fusion proteins was verified with doxycycline (tetracycline derivative) (Fig. 10). Biotinylated/tagged proteins were immunoprecipitated and identified by mass spectrometry (Fig. 11). As accepted, known ZRSR2 binding partners including members of the SF3B complex, U2AF1, U2AF2 and SRPK1 were identified, and their interaction with ZRSR2 was verified (Figs. 12-13). Novel partners that we verified include C1QBP and CCDC97 (Fig. 14). C1QBP is involved in metabolism and tumorigenesis and is known to associate with splicing factor SRSF1 and inhibit its binding to RNA. CCDC97 associates SF3B members, however its function unknown. We hypothesized that CCDC97 might regulate ZRSR2 expression and/or affect its splicing activity. We found that knockdown of CCDC97 (shRNA) does not alter ZRSR2 expression (Fig. 15). qPCR analysis showed changes in splicing of specific U2- and U12-introns and RNA-seq analysis further identified global splicing changes in CCDC97 knockdown TF1 cells (Figs. 16-17). Knockdown of CCDC97 impaired growth and colony-forming ability of TF1 cells (Fig. 18). We are currently studying further the affects of ZRSR2-CCDC97 interaction on splicing and downstream cellular pathways.

2D. High-throughput drug screen. To identify genes and pathways that synergize with ZRSR2 loss, we performed synthetic lethal screens using *ZRSR2*-deficient and control TF1 and K562 cells. For these experiments we used two parallel screening platforms: 1) A library of siRNAs directed against the entire tyrosine kinome (91 family members) as well as N-RAS, K-RAS, and control scrambled siRNAs, 2) a small-molecule panel (~130 drugs) of FDA-approved or late-stage development compounds. The screens identified several potentially interesting sensitivities in *ZRSR2* knockdown cell lines. For example, the knockdown cells exhibited increased sensitivity to siRNA targeting ERBB3/4, and to inhibitors of the MAPK pathway (RAF265, an inhibitor of WT and oncogenic RAF, and CI-1040, a MEK1/2 inhibitor) (Fig. 19). Encouragingly, our pathway analysis of mis-spliced genes in association with *ZRSR2* mutations in MDS patients showed enrichment of genes in the MAPK and ERBB signaling pathways. Interestingly, a germline *ERBB3* variant was identified as a candidate for predisposition to erythroid MDS/erythroleukemia (Braunstein et al., Leukemia. 2016). Our results suggest strong dependencies of *ZRSR2*-deficient cells on MAPK and ERBB pathways for their viability; and, importantly, MAPK and ERBB inhibitors may be effective in patient with *ZRSR2* mutations.

2E. ZRSR2 knockdown induces nonsense-mediated mRNA decay. Intron retention events induced by ZRSR2 mutations frequently introduce premature stop codons which can lead to either truncated products or induce nonsense-mediated mRNA decay (NMD). To test whether loss of ZRSR2 induces NMD, we used cycloheximide to inhibit NMD in ZRSR2 knockdown TF1 and K562 cells and measured and measured the effect on specific abnormal intron retention events (e.g. MAPK1 intron 2, MAPK9 intron 7). Cycloheximide treatment resulted in

increased expression of some of transcripts with retained introns, suggesting that these transcripts are degraded by NMD (Fig. 20)

Specific Aim 3: Investigate the role of ZRSR2 in myeloid transformation using *Zrsr2/Tet2*-deficent murine model.

3A. *Zrsr2*/*Tet2* double KO mice. In MDS, mutations of *ZRSR2* frequently co-occur with loss-of function mutations of *TET2*. To understand the cooperativity between these two genetic aberrations in the pathogenesis of myeloid disorders, we are generating *Zrsr2*/*Tet2* double KO mice. We will assess hematopoiesis and development of leukemia in the in the double KO mice.

3B. *ZRSR2* + *TET2* mutations in MDS samples. To assess gene expression and splicing changes caused by combined loss of *TET2* and *ZRSR2* in human cells, we performed RNA-seq on MDS bone marrow harboring either *ZRSR2* mutation alone, *TET2* mutation alone or *ZRSR2* + *TET2* mutations. We are currently analyzing the data using our Mis-Splicing Index.

3C. Combined silencing of *ZRSR2* and *ZRSR1*. The *ZRSR2* gene has a closely related paralog, *ZRSR1*; human *ZRSR1* is a pseudogene of *ZRSR2*, murine *Zrsr1* is a protein coding gene (Fig. 21). We reasoned that *Zrsr1* likely compensates for loss of *Zrsr2*. We are now studying the effects of loss of both *Zrsr1* and *Zrsr2* on hematopoiesis and splicing in myeloid cells. We generated murine myeloid 32D cells deficient in both ZRSR1 (RFP+) and ZRSR2 (GFP+). However, knockdown efficiency was low and those cells were outcompeted (Figs. 22-23). We then silenced *Zrsr1* in *Zrsr2* KO progenitors (Lin^-Kit^+). Colony assays and in vitro competitive proliferation assays showed that cells deficient for both genes formed fewer colonies and had a proliferative disadvantage over cells with knockdown of *Zrsr2* alone (Fig. 24). RNA-seq analysis showed that silencing *Zrsr1* in *Zrsr2* KO myeloid cells further affected splicing of U12-introns compared to *Zrsr2* KO cells withour silencing of *Zrsr1* (Figs. 25-26). We are now deleting *Zrsr2* gene in 32D cells using CRISPR/Cas9 (Fig. 27). We will then use *Zrsr1* shRNA to generate cells with knockdown of both *Zrsr1* and *Zrsr2*. In future, we will generate *Zrsr1/Zrsr2* double KO mice and study the mice for hematopoiesis and development of myelopoiesis.

- **Training and professional development opportunities**

Nothing to Report

- **Results dissemination**

Nothing to Report

- **Plans for the next reporting period**

The period of performance has ended. We will submit a final report.

4. IMPACT

- **Impact on the development of the principal discipline(s) of the project**

Our work helps clarify the mechanistic basis for the pervasive nature of splicing disruption in MDS, leading to an overall better understanding of the disease. Over the longer term, this study may help guide design of novel therapeutic strategies that will improve the health and lives of patients with MDS and related diseases.

- **Impact on other disciplines**

Nothing to Report

- **Impact on technology transfer**

Nothing to Report

- **Impact on society beyond science and technology?**

Knowledge gained from this study may be harnessed to develop rational, effective treatments for MDS patients with splicing mutations.

5. CHANGES/PROBLEMS

- **Changes that had a significant impact on expenditures**

Nothing to Report

- **Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents.**

Nothing to Report

6. PRODUCTS

Nothing to Report

7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

- **Individuals**

H. Phillip Koeffler, PI “no change”
Sigal Gery, Research Scientist “no change”
Li Chen, Post-Doc “no change”

- **Change in support**

Nothing to Report

- **Organizations**

Nothing to Report

8. SPECIAL RERPORTING REQUIREMENTS

None

Distinct and convergent consequences of splice factor mutations in myelodysplastic syndromes

Abstract

Myelodysplastic syndromes (MDS) are characterized by recurrent somatic alterations often affecting components of RNA splicing machinery. Mutations of splice factors SF3B1, SRSF2, ZRSR2 and U2AF1 occur in >50% of MDS. To assess the impact of spliceosome mutations on splicing and to identify common pathways/genes affected by distinct mutations, we performed RNA-sequencing of 32 MDS bone marrow samples harboring solely spliceosome mutations (including SF3B1, SRSF2 P95 or small deletions, U2AF1 S34 or Q157 hotspot mutations and ZRSR2 truncating alterations), but devoid of other co-occurring mutations. We uncover the landscape of splicing alterations in each splice factor mutant MDS and demonstrate that SRSF2 deletions cause highest number of splicing alterations compared with other spliceosome mutations. Although the mis-spliced events observed in different splice factor mutations were largely non-overlapping, a subset of genes, including *EZH2*, were aberrantly spliced in multiple mutant groups. Pathway analysis revealed dysregulated biological processes including RNA splicing and transport as well as several signaling cascades, which were observed repeatedly in different mutant groups, suggesting converging biological consequences downstream of distinct spliceosome mutations.

Introduction

Myelodysplastic syndromes (MDS) are a heterogeneous group of hematological disorders characterized by impaired myeloid differentiation, dysplasia, peripheral cytopenia and anemia. MDS is associated with aging and about 1/3rd of the cases progress to acute myeloid leukemia (AML)[1,2]. Clinically, MDS is categorized into several distinct subtypes based on percentage of blasts, lineage affected, presence of ring sideroblasts and karyotype.[3]

Recurring genetic abnormalities are observed in a majority of MDS patients, with somatic mutations of spliceosome genes being most prevalent, affecting about 60% of cases[4,5], suggesting that aberrant spliceosome function is a key determinant in pathogenesis of MDS. The spliceosome genes frequently mutated in MDS include SF3B1, SRSF2, U2AF1 and ZRSR2; and proteins encoded by each are involved in recognition of 3' splice site[6], suggesting a common biological outcome. Mutations of spliceosome genes are early genetic events and occur mutually exclusive of each other[5-7]. Functional studies have characterized splicing signatures associated with each spliceosome mutation and identified mis-spliced genes downstream of each mutation[8-18]. However, splicing changes were investigated usually in diverse cellular background with each study focusing on single splice factor. Murine models of common mutant alleles (U2AF1-S34F, SRSF2-P95H, SF3B1-K700E) have enabled elucidation of hematopoietic and splicing defects caused by these splice factor mutations[10,12-15], however, intronic sequences and exonic regulatory elements in human and mouse are poorly conserved. Therefore, mis-spliced genes identified in mouse models may not recapitulate the changes observed in clinical samples carrying these mutations. Therefore, the conundrum whether spliceosome mutations affect a

shared set of transcripts or biological processes in MDS needs to be carefully evaluated.

To address the functional outcome of different spliceosome mutations, we analyzed RNA splicing in MDS bone marrow samples harboring mutations of four different spliceosome genes in a single cohort. Our collective in-depth analyses of dysfunctional splicing in different mutant group uncovers a subset of genes aberrantly spliced in more than one mutant group. We also identified common biological pathways including RNA processing and splicing which are affected in different splice factor mutant MDS, indicating functionally convergent outcomes of distinct mutations.

Methods

RNA extraction and sequencing of MDS bone marrow samples

Bone marrow aspirates of MDS were obtained from the MLL Munich Leukemia Laboratory. Informed consent was obtained in accordance with the Declaration of Helsinki and approved by the Institutional Review Board of the MLL.

RNA was extracted from bone marrow mononuclear cells using Qiagen RNeasy Mini Kit (Qiagen). cDNA libraries were prepared using TruSeq RNA Sample Preparation Kit (Illumina) according to the manufacturer's protocol. polyA-selected RNA was used for library preparation and paired end adapters were ligated to DNA fragments prior to amplification and sequencing on HiSeq 4000.

Differential gene expression analysis

RNA-seq reads were mapped to mouse hg19 genome using STAR aligner.[19] Mapped reads of each gene were used to identify gene expression by featureCounts.[20] Gene expression counts were further normalized among samples based on the total numbers of all mapped reads. Differentially Expressed Genes (DEGs) were then identified from the normalized gene-level counts using the cut-off values for fold-change of 2 and adjusted p value of 0.05.

Splicing analysis

For the splicing analysis, we used our previously established pipeline.[16] Briefly, we utilized the difference in Mis-splicing Index (ΔMSI) to evaluate the variance in splicing efficiency between spliceosome mutant and WT MDS samples and used the ΔMSI values between the replicates to estimate the background noise. We further classified all the altered splice events as either intron retention, cryptic splice site usage (including alternative 5'/3' splice sites) or altered exon inclusion/exclusion. The RNA-seq junction reads at genomic and transcriptome locations were applied for characterization of the ratio of the productive transcript expression to the alternatively spliced expression. New amino acid sequences and premature termination codon sites of discovered splice events were predicted according to transcript annotation.

Pathway enrichment analyses

DAVID functional annotation analysis (https://david.ncifcrf.gov/) was utilized to identify enriched KEGG pathways in the DEGs as well as differentially spliced genes. Gene Set Enrichment Analysis (GSEA)[21] was used to identify enriched Hallmark gene

sets in the database MSigDB 6.1 comparing gene expression profiling between MDS samples with spliceosome mutations and those without mutations of splice factors.

RTPCR analysis for splicing of XBP1 and EZH2

Transcript levels of unspliced and spliced *XBP1* were estimated using RT-PCR. RNA was treated with DNase I and reverse transcribed using Maxima First Strand cDNA Synthesis Kit (Thermo). cDNA was amplified using primers CCTGGTTGCTGAAGAGGAGG and CCATGGGGAGATGTTCTGGAG and resolved on 3% agarose gel. Intensity of PCR bands corresponding to unspliced (145 bp) and spliced XBP1 (119 bp) were determined using ImageLab software (Biorad).

Aberrant splicing of EZH2 was assessed using quantitative RT-PCR with two primer pairs: TTTGGACAGAAGATTCATTGAATG and CGTTTTGGTGGGGTCTTTATC (PCR 1) and CCAACACTTATAAGCGGAAGAAC and AAATCCAACAGGCAATATATACCC (PCR 2).

Results

Global profiling of mis-splicing in different spliceosome mutant MDS

To identify splicing changes caused by different spliceosome mutations, we selected MDS samples harboring mutations of either *SF3B1*, *SRSF2*, *U2AF1* or *ZRSR2* (Table 1). To preclude the effect of other genetic alterations on RNA splicing, samples were carefully chosen to avoid accompanying mutation in other commonly mutated genes.

We screened a panel of 36 genes usually altered in myeloid diseases (Supplementary Table 1) using targeted sequencing and selected samples with a lone mutation in one of the four splice factor gene. For SRSF2, apart from the P95 hotspot mutation, we also selected rare samples with small deletions in *SRSF2* gene (SRSF2 del) (Table 1). In total, we performed RNA-sequencing on poly-A selected RNA from 24 MDS bone marrow with spliceosome mutation and eight samples without spliceosome mutation as control.

We quantified splicing defects that occur in different mutant groups using the approach described previously[16]. Global splicing alterations were profiled by subtracting the average mis-splicing index (MSI) of the eight MDS samples without spliceosome mutations (splice factor wild-type MDS: SF WT MDS) to obtain the ΔMSI value for every splicing event in each spliceosome mutant sample. These analyses were performed for different splicing events categorized as aberrant intron retention, altered exon inclusion/exclusion and cryptic splice site usage. First, our global profiling using principal component analysis (PCA) resulted in segregation of samples belonging to different mutant groups in one or more splicing categories (Figure 1A-C). For instance, SF3B1 mutant samples clustered together in our analysis of cryptic splice site usage distinct from other mutant samples (Figure 1B), while intron retention profiling segregated ZRSR2 mutant samples from other spliceosome mutant MDS samples (Figure 1A). Interestingly, MDS samples with SRSF2 del were clustered together in all three mis-splicing classes, separated not only from samples with other splice factor mutations but also from the SRSF2 P95 hotspot mutants (Figure 1A-C). On the contrary, U2AF1 mutant samples did not cluster in any of our analyses, indicating they do not exhibit specific and robust mis-splicing changes (Figure 1A-C). However, when we checked differential splicing in U2AF1 mutant samples compared only to the SF

WT MDS, we obtained two distinct clusters which segregated them according to mutation sites – S34 and R156/Q157 (referred to as U2AF1 Q157 mut) (Supplementary Figure 1), indicating distinct splicing consequences of these two hotspots, similar to the findings in previous reports.[8,9]

Next, we determined differential splicing events for each mutation group with 10% as the cutoff value for ΔMSI. We further required significant change in splicing (Fisher's exact test; $p<0.01$) in at least 75% pairwise comparisons (mutant vs SF WT MDS). As shown in Figure 2A-C, we observed distinct clusters of differentially mis-spliced events in each mutant group, except for U2AF1 mutant samples, which showed largely non-overlapping splicing changes within the group (Figure 2A-C).

This analyses also revealed an overall low overlap of mis-spliced events downstream of different spliceosome mutations in MDS (Figure 2A-C and Supplementary Table 2). Although all spliceosome mutant MDS bone marrow exhibited significantly higher number of mis-spliced events compared to the SF WT MDS samples, SRSF2 deletion associated with the overall largest number of mis-splicing events amongst all groups (Figure 2D and Supplementary Figure 2A). Samples with mutations of either ZRSR2, SF3B1 or SRSF2 del exhibited consistent mis-splicing events across pairwise comparisons, while splicing changes downstream of U2AF1 S34, U2AF1 Q157 or SRSF2 P95 mutations showed variability in different samples (Supplementary Figure 2B). We also analyzed the proportion of productive transcript expression for genes aberrantly spliced in different splice factor mutant samples. Although we detected similar total expression for these genes in the mutant and SF WT samples, the levels of productive transcripts were reduced by 30-40% in mutant samples compared to SF WT samples, irrespective of the mutation type (Figure 2E).

Mis-splicing signatures in different splice factor mutant samples

Our analysis showed that SF3B1 mutant samples exhibited predominantly increased usage of non-canonical splice sites (Figure 2D), which is congruent with the published results.[11-13,22] Amongst events with usage of abnormal splice sites, splicing to cryptic 3' splice site was prevalent only in SF3B1 mutant samples compared with other splice factor mutant MDS (Figure 3A). Increased use of cryptic splice sites in cells harboring SF3B1 mutation was possibly caused due to weaker 3' splice sites in SF3B1 mutant samples compared to other mutant groups (Supplementary Figure 3A).

Aberrant intron retention was the major splicing defect in ZRSR2 mutant samples (Figure 2D). Impaired splicing affecting primarily the U12-type introns, similar to the pattern we have described before[16], was exclusive to ZRSR2 mutant MDS, while other splice factor mutations did not display any propensity towards selective mis-splicing of U2-type vs U12-type introns (Figure 3B). Intron retention events resulted invariably in transcripts with premature termination codon in the three mutant groups (SRSF2 del, U2AF1 S34 and ZRSR2 mut) with substantial number of retained introns (Supplementary Figure 3B).

In accordance with previous reports,[14,15,23] SRSF2 P95 missense mutation resulted in enrichment for the CCNG motif in exons preferentially included and for GGNG motif in preferentially excluded exons (Figure 3C). Interestingly, two samples with the in-frame deletion involving P95 residue, which showed a significantly higher rate of aberrant splicing than the P95 missense mutant MDS samples (Figure 2D), exhibited stronger enrichment for the CCNG motif in preferentially included exons. The enrichment for the GGNG motif in the repressed exons was similar in both P95 mutant and deleted MDS samples (Figure 3C). These observations indicate an essential role of P95

residue in recognition of exonic splicing enhancers and that its mutation or loss alters RNA binding ability of SRSF2 leading to enhanced inclusion of exons containing CCNG motif. In contrast, the non-P95 SRSF2 del sample (SRSF2 mutant 6) showed enhanced exclusion of exons with both CCNG and GGNG motifs, suggesting alternative consequences on exon recognition and RNA splicing. Overall, we also noted that there was a low overlap of significantly mis-spliced events between the P95 and SRSF2 del groups (Supplementary Figure 3C).

The two U2AF1 hotspot mutations, S34 and Q157, generally affected low number of largely non-overlapping splicing events within each group compared with other splice factor mutations (Figure 2A-D). However, as previously described,[8,24] the two U2AF1 mutations affected splicing of discreet subset of genes (Supplementary Figure 3D), indicating that both hotspot mutations trigger distinct downstream splicing events.

Genes affected by different spliceosome mutations are largely distinct

Prior studies investigating splicing changes caused by mutations of either SRSF2, SF3B1, U2AF1 or ZRSR2 using different in vivo and in vitro models have indicated distinct splicing signatures.[8-18,25] In the present study, non-overlapping nature of mis-splicing events was also evident in the different mutant groups (Supplementary Table 2). However, we identified that splicing of 115 genes was commonly altered in two or more spliceosome mutant groups (Figure 4A). This included increased expression of an *EZH2* isoform containing a 'poison' cassette exon in SRSF2 mutant (both P95 and deletion samples) and U2AF1 mutant samples (Figure 4B-C). Our quantitative RT-PCR analysis verified aberrant splicing to this atypical exon (located in intron 9) containing a premature stop codon, in SRSF2 and U2AF1 mutant samples (Figure 4D).

Surprisingly, some of the SF3B1 mutant MDS samples also had a trend towards increased mis-splicing of *EZH2*, suggesting loss of EZH2 expression is a common event in SRSF2, U2AF1 and SF3B1 mutant MDS. Other than *EZH2*, 10 other genes were mis-spliced in more than two mutant group (Figure 4A). These genes may represent possible common mediators downstream of spliceosome mutations.

U2AF1 S34 and Q157 mutations altered splicing of distinct sets of introns, with only a single event in *TAPT1* gene shared in two groups (Figure 4A and Supplementary Figure 3D). We observed that the frameshift spliced variant of XBP1 (sXBP1) occurred at higher levels in Q157 mutants compared to S34 mutants, suggesting increased ER stress and activation of unfolded protein response (UPR) pathway in Q157 mutant MDS (Supplementary Figure 4). Amongst other splice factor mutations, SRSF2 P95 mutant samples also exhibited increased splicing of XBP1 (Supplementary Figure 4).

Pathways affected by mis-splicing in different spliceosome mutations

To gain further insights into the functional consequences of aberrant splicing in spliceosome mutant MDS, we examined biological pathways commonly enriched in significantly mis-spliced genes in different mutant groups. RNA splicing was the top pathway significantly dysregulated in SRSF2 mutant (both P95 and del samples), U2AF1 S34 and ZRSR2 mutant groups (Figure 5). In addition, RNA processing and transport were also affected by aberrant splicing in multiple mutant groups (Figure 5), suggesting dysregulation of RNA processes as one of the common biological consequence of spliceosome mutation. Aberrant splicing of components of different signaling networks in two or more mutant groups including AMPK signaling, NOD-like receptor signaling, TNF-α signaling, MAPK signaling and EGFR signaling was also evident (Figure 5). Genes encoding for components of protein processing in the

endoplasmic reticulum were mis-spliced in SRSF2 P95, SRSF2 del and ZRSR2 mutant groups (Figure 5). In addition, heme metabolism, chronic myeloid leukemia and apoptosis were other pathways relevant to MDS which were altered in both SRSF2 and ZRSR2 mutant samples (Figure 5).

Gene expression profiling reveals dysregulated heme metabolism and cell cycle regulation in spliceosome mutant MDS

We further performed global gene expression profiling of MDS samples to ascertain if shared biological pathways are affected in different spliceosome mutant groups. Since the MDS samples in our cohort were devoid of other common mutations in epigenetic regulators and transcription factors, gene expression changes observed in our samples truly reflect the effect of spliceosome gene mutation, possibly through altered splicing of transcriptional regulators. GSEA revealed that genes involved in heme metabolism were differentially expressed in all the spliceosome mutant MDS groups compared with SF WT MDS (Figure 6 and Supplementary Table 3), possibly contributing to perturbed erythropoiesis. We also identified dysregulated expression of genes associated with cell cycle in different spliceosome mutant MDS groups, which resulted in enrichment of cell cycle related gene sets (E2F targets and G2M checkpoint) (Figure 6 and Supplementary Table 3).

Discussion

Highly recurrent mutations in genes encoding for splice factors provide compelling evidence of the role of dysregulated RNA splicing in pathogenesis of MDS. Initial studies to decipher the consequences of spliceosome mutations utilized diverse

models, both in vitro and in vivo, and primarily focused on a single mutation. Although these studies provided evidence of increased mis-splicing in spliceosome mutant MDS and recognized mis-splicing pattern specific to each mutation, use of different cell types - either human or murine - did not allow a definitive comparison of splicing changes downstream of different mutations. Therefore, the question whether mutually exclusive spliceosome mutations cause shared splicing changes remained unknown. Qiu et al attempted to address this conundrum using RNA-mediated oligonucleotide annealing, selection, and ligation coupled with next-generation sequencing of MDS samples.[26] Although only 5,500 splicing events were interrogated, the study concluded that MDS-linked splicing signatures are primarily involved in cell cycle control and DNA damage responses.[26] Recently, $CD34^+$ bone marrow cells from MDS patients carrying different spliceosome mutations were interrogated in a single cohort using RNA-sequencing. Although the impact of SRSF2, SF3B1 and U2AF1 mutations on splicing was distinctive, some common dysregulated pathways including RNA splicing, protein synthesis and mitochondrial dysfunction were observed.[27] More recently, splicing changes and transcriptional consequences of abnormal RNA splicing were also investigated in a large cohort of MDS samples.[24]

In accordance with these previous reports, our analyses defined largely distinct splicing patterns for the four frequently mutated splice factor genes, albeit a small subset of genes commonly mis-spliced in different groups was identified. Amongst these, we verified that splicing to a ‘poison exon’ of *EZH2* gene occurred in U2AF1, SRSF2 and SF3B1 mutant cases, highlighting suppression of EZH2 as one of the common consequences of splice factor mutations in MDS. *EZH2* encodes for the catalytic subunit of the polycomb repressive complex 2 (PRC2), which possesses H3K27 methyltransferase activity. Importantly, *EZH2* is located at chromosome 7q36,

a region frequently deleted (-7 and 7q-) in MDS. Inactivating mutations of *EZH2* are also observed in MDS, which are associated with its decreased expression level.[28-32] Importantly, deletion of *Ezh2* in mice results in MDS/MPN-like diseases,[28] thus substantiating the pathogenetic role of EZH2 deficiency in development of disease. However, scarce detection of aberrant splicing in other MDS/AML-related genes in splice factor mutant MDS and little overlap of mis-spliced events between different mutant groups suggests that multiple splicing events, rather than a single event, may collectively contribute to the pathogenesis of MDS.

In addition to the well-described missense mutation at proline 95 in *SRSF2* gene, we also analyzed MDS cases with small deletions; two cases with identical 24- nucleotide in-frame deletion including P95 and one case with frameshift deletion at arginine 150 residue. These rare deletions affecting *SRSF2* have been previously reported in MDS and CMML,[33-37] however, their consequences on splicing have not been investigated. We observe that these microdeletions of *SRSF2* caused a greater impact on aberrant splicing compared with the P95 missense mutation. While deletion of P95 residue alters exon recognition via preferential recognition of C-rich exonic splicing enhancer motif, similar to P95 missense mutation, non-P95 frameshift deletion does not show this specificity. This is in agreement with the described role of P95, which is located in RNA recognition motif of SRSF2, in sequence specific RNA binding to regulate alternative splicing.[14,38,39]

Despite largely non-overlapping mis-spliced events observed in MDS harboring different spliceosome mutations, the genes affected by altered splicing in different groups converged into several common biological pathways. These included RNA processing, transport and splicing, indicating dysregulation of RNA machinery as a common consequence of splice factor mutations in MDS. Recently, similar enrichment

of RNA processing and splicing genes in MDS carrying mutations of SF3B1, SRSF2 and U2AF1 was also reported.[27] Liang et al also observed that SRSF2 P95H mutation affected splicing of several members of the hnRNP and SR families of proteins and suggested a “splicing-cascade” phenotype wherein mutation of a single splicing factor leads to widespread modifications in multiple RNA processing and splicing proteins.[38] Other commonly dysregulated pathways that we identified included various signaling pathways, disruption of which possibly impedes growth and differentiation of myeloid cells. For instance, mitogen-activated protein kinase (MAPK) signaling is vital to self-renewal and differentiation as well as cell survival and apoptosis.[40,41] In addition, splicing of genes encoding for members of AMPK signaling pathway is altered in SF3B1, SRSF2 and ZRSR2 mutant MDS. This signaling pathway displays tumor suppressor activity in AML through the repression of mTOR-dependent translation. [42,43]. However, sirtuin signaling pathway, which was significantly dysregulated across the mis-spliced genes in SF3B1, SRSF2 and U2AF1 mutant MDS samples in a recent study,[27] was not enriched in any of the splice factor mutant groups of our cohort.

Overall, this study profiles splicing alterations and identifies targets of each spliceosome mutation in MDS. It also illustrates that despite distinctive consequences of different splice factors mutations on RNA splicing, genes affected by aberrant splicing converge in common dysregulated pathways and processes, many of which can be either directly related to disease pathophysiology or are prospective novel mediators of disease and need to be further investigated.

Authorship

V.M. conceived the study, designed and performed research, analysed data and wrote the manuscript; J.L. and S.Z. performed bioinformatics and statistical analyses and wrote the manuscript; T.W.W. and L.H. performed research and analysed data; M.M. and T.H. provided clinical samples and patient information; H.Y. performed and supervised bioinformatics and statistical analyses and wrote the manuscript; H.P.K. conceived and supervised the study, interpreted the data and wrote the manuscript. All authors reviewed and approved the manuscript.

Acknowledgements

We thank the Melamed Family and Reuben Yeroushalmi for their generous support. This work was funded by the Leukemia and Lymphoma Society, the Singapore Ministry of Health's National Medical Research Council (NMRC) under its Singapore Translational Research (STaR) Investigator Award to H. Phillip Koeffler (NMRC/STaR/0021/2014), the NMRC Centre Grant awarded to National University Cancer Institute of Singapore (NMRC/CG/012/2013) and the National Research Foundation Singapore and the Singapore Ministry of Education under its Research Centres of Excellence initiatives.

This research is also supported by the RNA Biology Center at the Cancer Science Institute of Singapore, NUS, as part of funding under the Singapore Ministry of Education's Tier 3 grants, grant number MOE2014-T3-1-006

Conflict-of-interest disclosure

Dr. Haferlach declares part ownership of MLL Munich Leukemia Laboratory. Dr. Meggendorfer is an employee of MLL Munich Leukemia Laboratory. The remaining authors declare no competing financial interests.

References

1. Ma X. Epidemiology of myelodysplastic syndromes. *Am J Med*. 2012;125(7 Suppl):S2-5.
2. Chamseddine AN, Jabbour E, Kantarjian HM, Bohannan ZS, Garcia-Manero G. Unraveling Myelodysplastic Syndromes: Current Knowledge and Future Directions. *Curr Oncol Rep*. 2016;18(1):4.
3. Steensma DP. Myelodysplastic Syndromes: Diagnosis and Treatment. *Mayo Clin Proc*. 2015;90(7):969-983.
4. Haferlach T, Nagata Y, Grossmann V, et al. Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. *Leukemia*. 2014;28(2):241-247.
5. Papaemmanuil E, Gerstung M, Malcovati L, et al. Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood*. 2013;122(22):3616-3627; quiz 3699.
6. Yoshida K, Sanada M, Shiraishi Y, et al. Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature*. 2011;478(7367):64-69.
7. Damm F, Kosmider O, Gelsi-Boyer V, et al. Mutations affecting mRNA splicing define distinct clinical phenotypes and correlate with patient outcome in myelodysplastic syndromes. *Blood*. 2012;119(14):3211-3218.
8. Ilagan JO, Ramakrishnan A, Hayes B, et al. U2AF1 mutations alter splice site recognition in hematological malignancies. *Genome Res*. 2015;25(1):14-26.
9. Okeyo-Owuor T, White BS, Chatrikhi R, et al. U2AF1 mutations alter sequence specificity of pre-mRNA binding and splicing. *Leukemia*. 2015;29(4):909-917.
10. Shirai CL, Ley JN, White BS, et al. Mutant U2AF1 Expression Alters Hematopoiesis and Pre-mRNA Splicing In Vivo. *Cancer Cell*. 2015;27(5):631-643.
11. Alsafadi S, Houy A, Battistella A, et al. Cancer-associated SF3B1 mutations affect alternative splicing by promoting alternative branchpoint usage. *Nat Commun*. 2016;7:10615.
12. Mupo A, Seiler M, Sathiaseelan V, et al. Hemopoietic-specific Sf3b1-K700E knock-in mice display the splicing defect seen in human MDS but develop anemia without ring sideroblasts. *Leukemia*. 2017;31(3):720-727.
13. Obeng EA, Chappell RJ, Seiler M, et al. Physiologic Expression of Sf3b1(K700E) Causes Impaired Erythropoiesis, Aberrant Splicing, and Sensitivity to Therapeutic Spliceosome Modulation. *Cancer Cell*. 2016;30(3):404-417.
14. Kim E, Ilagan JO, Liang Y, et al. SRSF2 Mutations Contribute to Myelodysplasia by Mutant-Specific Effects on Exon Recognition. *Cancer Cell*. 2015;27(5):617-630.
15. Komeno Y, Huang YJ, Qiu J, et al. SRSF2 Is Essential for Hematopoiesis, and Its Myelodysplastic Syndrome-Related Mutations Dysregulate Alternative Pre-mRNA Splicing. *Mol Cell Biol*. 2015;35(17):3071-3082.
16. Madan V, Kanojia D, Li J, et al. Aberrant splicing of U12-type introns is the hallmark of ZRSR2 mutant myelodysplastic syndrome. *Nat Commun*. 2015;6:6042.
17. Darman RB, Seiler M, Agrawal AA, et al. Cancer-Associated SF3B1 Hotspot Mutations Induce Cryptic 3' Splice Site Selection through Use of a Different Branch Point. *Cell Rep*. 2015;13(5):1033-1045.
18. Zhang J, Lieu YK, Ali AM, et al. Disease-associated mutation in SRSF2 misregulates splicing by altering RNA-binding affinities. *Proc Natl Acad Sci U S A*. 2015;112(34):E4726-4734.
19. Dobin A, Gingeras TR. Mapping RNA-seq Reads with STAR. *Curr Protoc Bioinformatics*. 2015;51:11 14 11-19.
20. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30(7):923-930.
21. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*. 2005;102(43):15545-15550.

22. Dolatshad H, Pellagatti A, Fernandez-Mercado M, et al. Disruption of SF3B1 results in deregulated expression and splicing of key genes and pathways in myelodysplastic syndrome hematopoietic stem and progenitor cells. *Leukemia*. 2015;29(5):1092-1103.
23. Kon A, Yamazaki S, Nannya Y, et al. Physiological Srsf2 P95H expression causes impaired hematopoietic stem cell functions and aberrant RNA splicing in mice. *Blood*. 2018;131(6):621-635.
24. Shiozawa Y, Malcovati L, Galli A, et al. Aberrant splicing and defective mRNA production induced by somatic spliceosome mutations in myelodysplasia. *Nat Commun*. 2018;9(1):3649.
25. Przychodzen B, Jerez A, Guinta K, et al. Patterns of missplicing due to somatic U2AF1 mutations in myeloid neoplasms. *Blood*. 2013;122(6):999-1006.
26. Qiu J, Zhou B, Thol F, et al. Distinct splicing signatures affect converged pathways in myelodysplastic syndrome patients carrying mutations in different splicing regulators. *RNA*. 2016;22(10):1535-1549.
27. Pellagatti A, Armstrong RN, Steeples V, et al. Impact of spliceosome mutations on RNA splicing in myelodysplasia: dysregulated genes/pathways and clinical associations. *Blood*. 2018;132(12):1225-1240.
28. Muto T, Sashida G, Oshima M, et al. Concurrent loss of Ezh2 and Tet2 cooperates in the pathogenesis of myelodysplastic disorders. *J Exp Med*. 2013;210(12):2627-2639.
29. Ernst T, Chase AJ, Score J, et al. Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. *Nat Genet*. 2010;42(8):722-726.
30. Nikoloski G, Langemeijer SM, Kuiper RP, et al. Somatic mutations of the histone methyltransferase gene EZH2 in myelodysplastic syndromes. *Nat Genet*. 2010;42(8):665-667.
31. Rinke J, Muller JP, Blaess MF, et al. Molecular characterization of EZH2 mutant patients with myelodysplastic/myeloproliferative neoplasms. *Leukemia*. 2017;31(9):1936-1943.
32. Shih AH, Abdel-Wahab O, Patel JP, Levine RL. The role of mutations in epigenetic regulators in myeloid malignancies. *Nat Rev Cancer*. 2012;12(9):599-612.
33. Wu SJ, Kuo YY, Hou HA, et al. The clinical implication of SRSF2 mutation in patients with myelodysplastic syndrome and its stability during disease evolution. *Blood*. 2012;120(15):3106-3111.
34. Thol F, Kade S, Schlarmann C, et al. Frequency and prognostic impact of mutations in SRSF2, U2AF1, and ZRSR2 in patients with myelodysplastic syndromes. *Blood*. 2012;119(15):3578-3584.
35. Makishima H, Visconte V, Sakaguchi H, et al. Mutations in the spliceosome machinery, a novel and ubiquitous pathway in leukemogenesis. *Blood*. 2012;119(14):3203-3210.
36. Hirabayashi S, Flotho C, Moetter J, et al. Spliceosomal gene aberrations are rare, coexist with oncogenic mutations, and are unlikely to exert a driver effect in childhood MDS and JMML. *Blood*. 2012;119(11):e96-99.
37. Meggendorfer M, Roller A, Haferlach T, et al. SRSF2 mutations in 275 cases with chronic myelomonocytic leukemia (CMML). *Blood*. 2012;120(15):3080-3088.
38. Liang Y, Tebaldi T, Rejeski K, et al. SRSF2 mutations drive oncogenesis by activating a global program of aberrant alternative splicing in hematopoietic cells. *Leukemia*. 2018.
39. Bourgeois CF, Lejeune F, Stevenin J. Broad specificity of SR (serine/arginine) proteins in the regulation of alternative splicing of pre-messenger RNA. *Prog Nucleic Acid Res Mol Biol*. 2004;78:37-88.
40. Chung E, Kondo M. Role of Ras/Raf/MEK/ERK signaling in physiological hematopoiesis and leukemia development. *Immunol Res*. 2011;49(1-3):248-268.
41. Miranda MB, Johnson DE. Signal transduction pathways that contribute to myeloid differentiation. *Leukemia*. 2007;21(7):1363-1377.
42. Green AS, Chapuis N, Lacombe C, Mayeux P, Bouscary D, Tamburini J. LKB1/AMPK/mTOR signaling pathway in hematological malignancies: from metabolism to cancer cell biology. *Cell Cycle*. 2011;10(13):2115-2120.
43. Green AS, Chapuis N, Maciel TT, et al. The LKB1/AMPK signaling pathway has tumor suppressor activity in acute myeloid leukemia through the repression of mTOR-dependent oncogenic mRNA translation. *Blood*. 2010;116(20):4262-4273.

Figure Legends

Figure 1. Principle component analysis (PCA) based on global profiling of aberrant splicing in spliceosome mutant MDS samples. (**A-C**) Splicing events were categorized as intron retention (A), cryptic splice site usage (B) and altered exon inclusion/exclusion (C) for PCA analysis.

Figure 2. Clustering of mis-spliced events in spliceosome mutant MDS compared to SF WT MDS. (**A-C**) Heat maps show cluster of significantly mis-spliced events for each mutant group. Splicing events were categorized as intron retention (A), cryptic splice site usage (B) and altered exon inclusion/exclusion (C). (**D**) Number and type of significant splicing changes identified in each group bearing a single spliceosome mutation. (**E**) Fold change in average expression of total and productive transcripts in mutant samples compared to SF WT MDS samples for significantly mis-spliced genes.

Figure 3. Splicing signatures for different splice factor mutant MDS samples. (**A**) Proportion of 5' and 3' cryptic splice sites in different spliceosome mutant MDS samples. (**B**) Frequency of U12-type intron-containing genes aberrantly spliced in MDS samples harboring mutations of spliceosome genes. (**C**) Relative frequency of CCNG and GGNC motifs in differentially spliced exons (significantly included or excluded) in individual SRSF2 mutant MDS samples.

Figure 4. Aberrant splicing of *EZH2* is a common event in different spliceosome mutant MDS samples. (**A**) Genes mis-spliced in two or more mutant groups. Genes aberrantly spliced in >2 spliceosome mutant MDS are highlighted in red. (**B**) Representative IGV snapshot depicts normalized RNA-Seq reads mapped to the genomic region encompassing exons 9-10 of *EZH2* locus in different spliceosome mutant MDS samples. (**C**) Proportion of RNA-seq reads supporting aberrant splicing of EZH2. (**D**) Experimental verification of aberrant splice junctions of *EZH2*. Quantitative RT-PCR analysis was performed using indicated primers to determine relative levels of mis-spliced transcripts in mutant and control samples.

Figure 5. Functional enrichment analysis based on significantly mis-spliced genes in spliceosome mutant MDS samples. Plots depict significant KEGG and Wiki pathways (p value<0.05) commonly enriched in genes aberrantly spliced in different spliceosome mutant MDS.

Figure 6. Gene expression profiling of spliceosome mutant MDS. GSEA plots of heme metabolism and cell cycle-related gene signatures (HALLMARK_E2F_TARGETS and HALLMARK_G2M_CHECKPOINT) in spliceosome mutant MDS. NES: Normalized enrichment score.

Figure 1

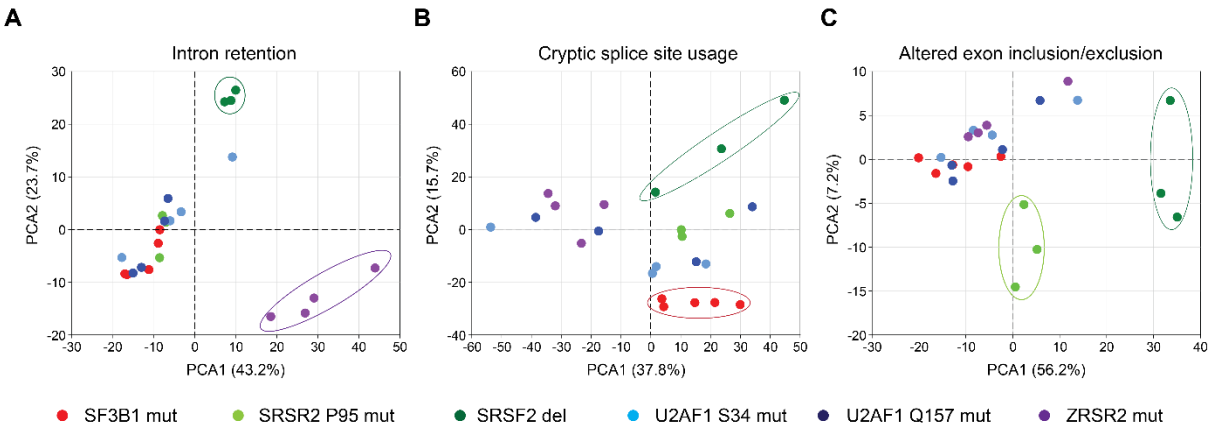

A
Intron retention
PCA2 (23.7%)
PCA1 (43.2%)
B
Cryptic splice site usage
PCA2 (15.7%)
PCA1 (37.8%)
C
Altered exon inclusion/exclusion
PCA2 (7.2%)
PCA1 (56.2%)
SF3B1 mut
SRSR2 P95 mut
SRSF2 del
U2AF1 S34 mut
U2AF1 Q157 mut
ZRSR2 mut

Figure 2

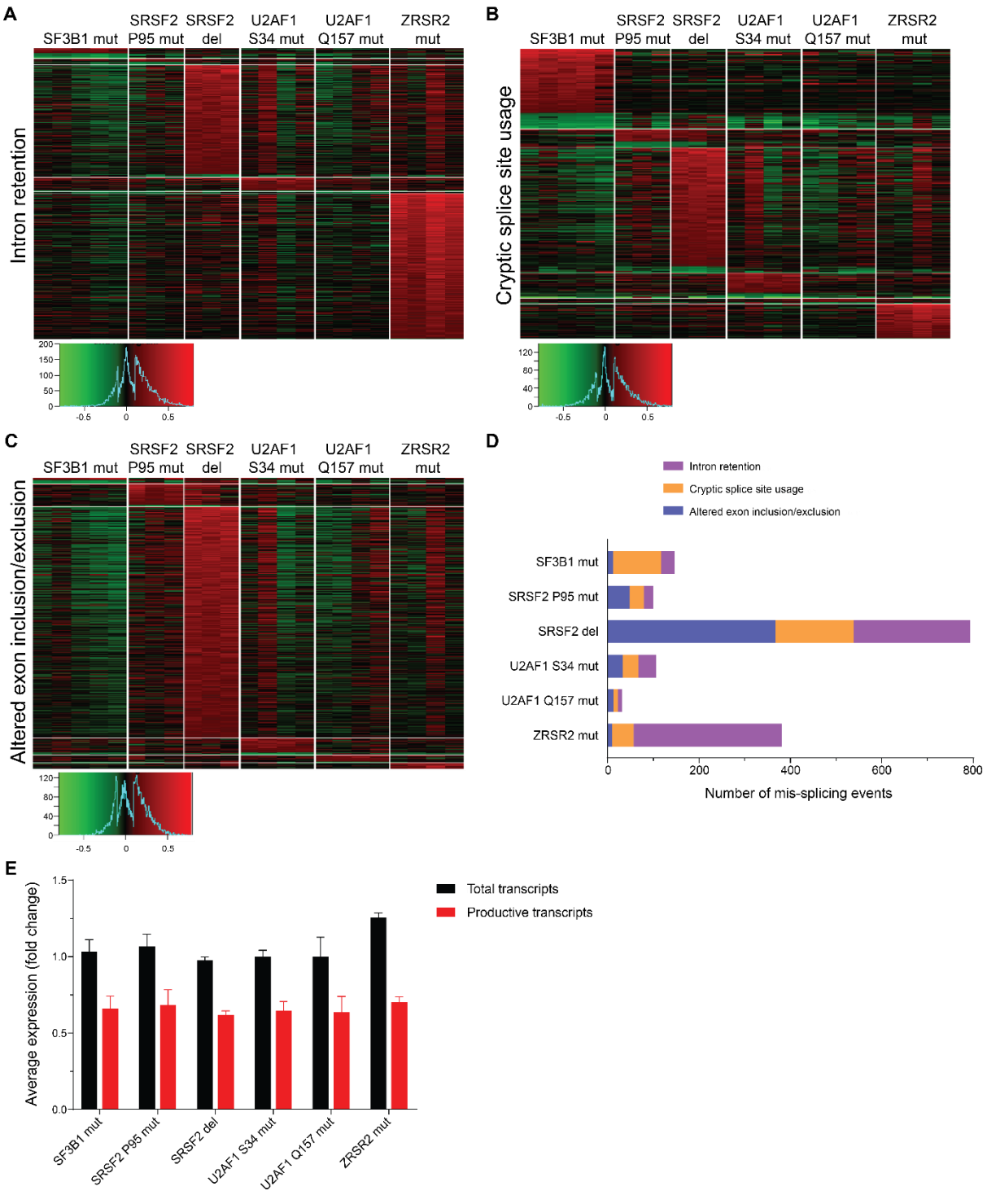

A
SF3B1 mut
SRSF2 P95 mut
SRSF2 del
U2AF1 S34 mut
U2AF1 Q157 mut
ZRSR2 mut
Intron retention
B
Cryptic splice site usage
C
Altered exon inclusion/exclusion
D
Intron retention
Cryptic splice site usage
Altered exon inclusion/exclusion
Number of mis-splicing events
E
Total transcripts
Productive transcripts
Average expression (fold change)

Figure 3

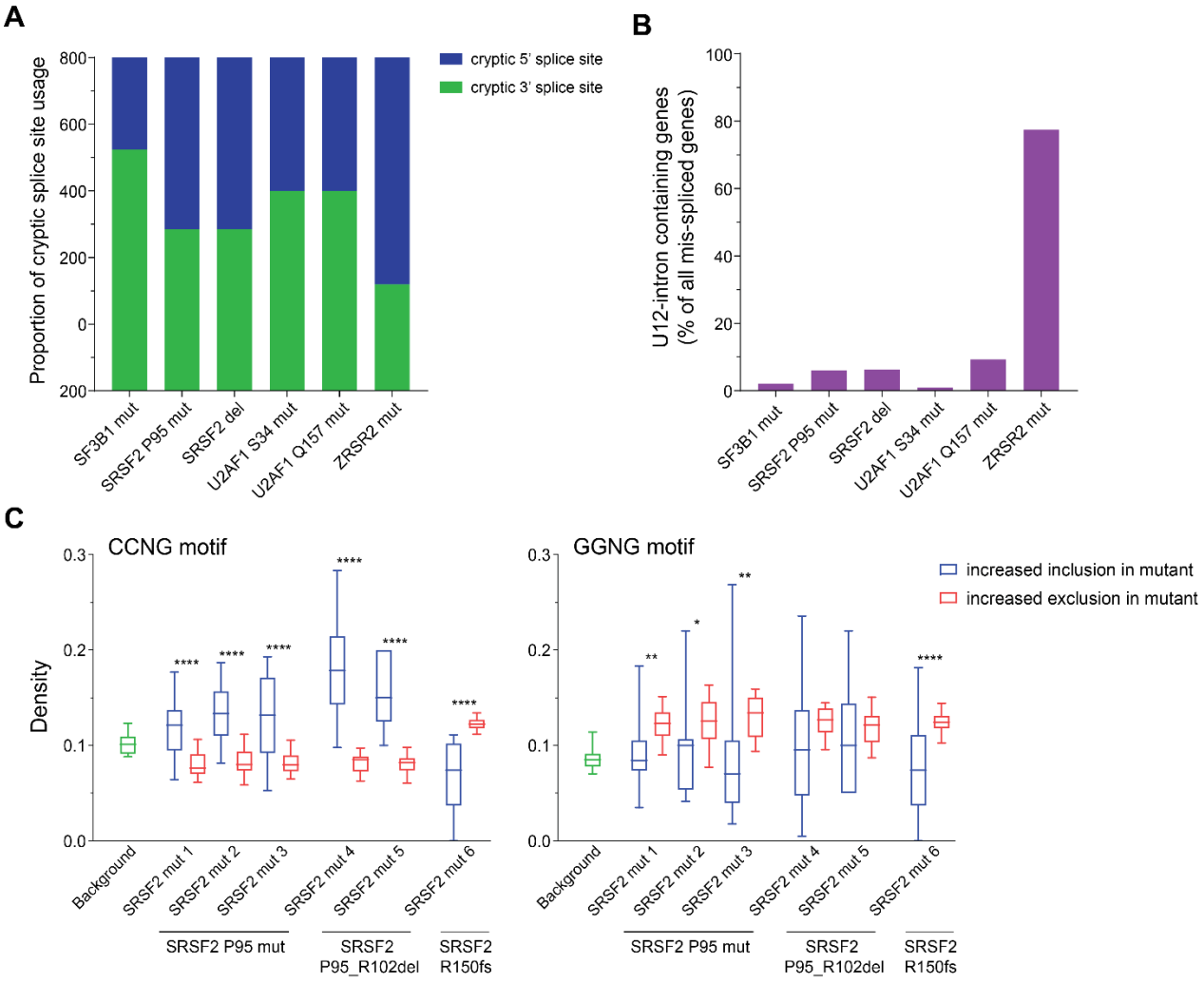

A
Proportion of cryptic splice site usage
800
600
400
200
0
200
cryptic 5' splice site
cryptic 3' splice site
SF3B1 mut
SRSF2 P95 mut
SRSF2 del
U2AF1 S34 mut
U2AF1 Q157 mut
ZRSR2 mut
B
U12-intron containing genes
(% of all mis-spliced genes)
100
80
60
40
20
0
SF3B1 mut
SRSF2 P95 mut
SRSF2 del
U2AF1 S34 mut
U2AF1 Q157 mut
ZRSR2 mut
C
CCNG motif
GGNG motif
Density
0.3
0.2
0.1
0.0
increased inclusion in mutant
increased exclusion in mutant
Background
SRSF2 mut 1
SRSF2 mut 2
SRSF2 mut 3
SRSF2 mut 4
SRSF2 mut 5
SRSF2 mut 6
SRSF2 P95 mut
SRSF2 P95_R102del
SRSF2 R150fs

Figure 4

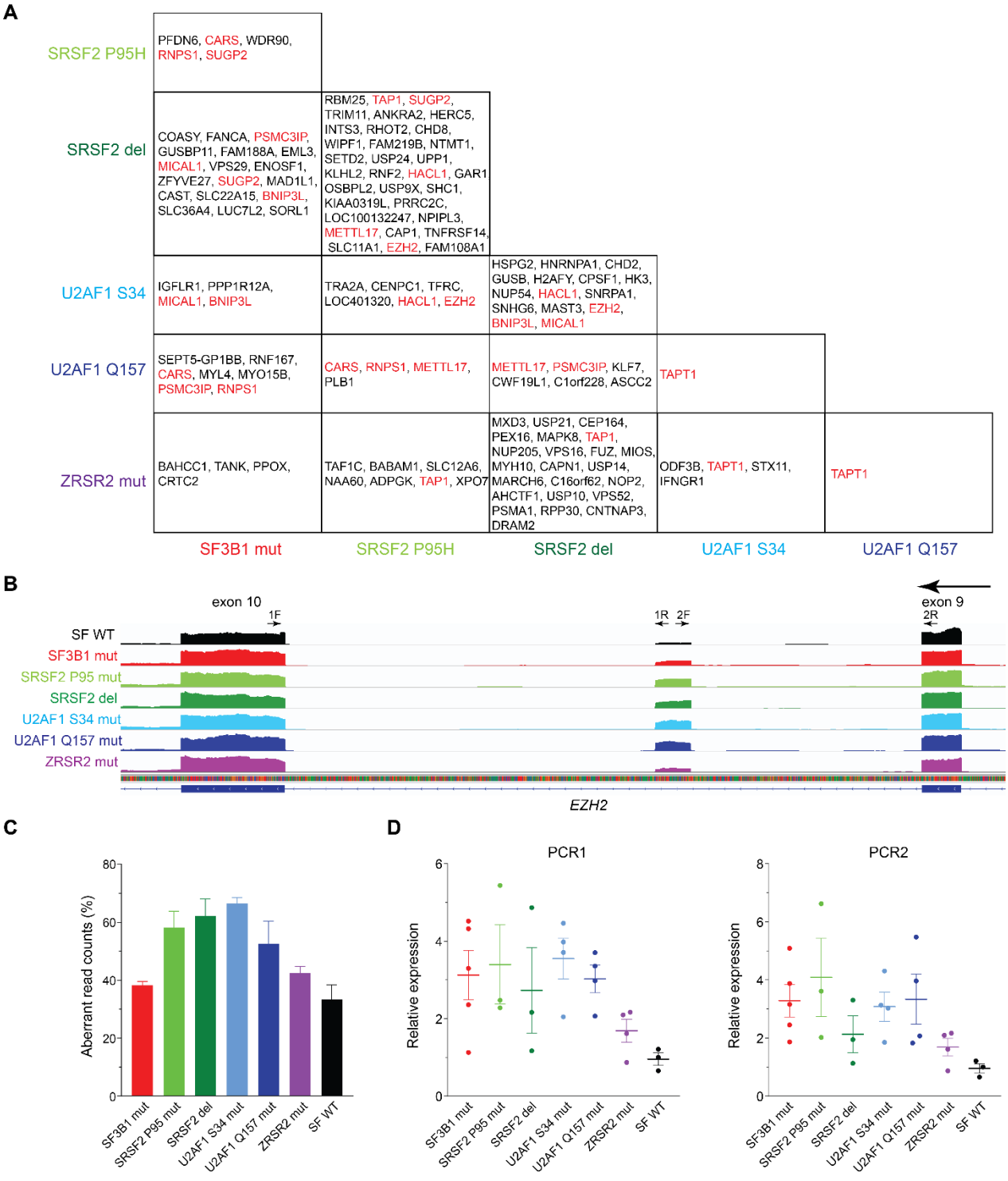

A
SRSF2 P95H
SRSF2 del
U2AF1 S34
U2AF1 Q157
ZRSR2 mut
PFDN6, CARS, WDR90, RNPS1, SUGP2
COASY, FANCA, PSMC3IP, GUSBP11, FAM188A, EML3, MICAL1, VPS29, ENOSF1, ZFYVE27, SUGP2, MAD1L1, CAST, SLC22A15, BNIP3L, SLC36A4, LUC7L2, SORL1
RBM25, TAP1, SUGP2, TRIM11, ANKRA2, HERC5, INTS3, RHOT2, CHD8, WIPF1, FAM219B, NTMT1, SETD2, USP24, UPP1, KLHL2, RNF2, HACL1, GAR1 OSBPL2, USP9X, SHC1, KIAA0319L, PRRC2C, LOC100132247, NPIPL3, METTL17, CAP1, TNFRSF14, SLC11A1, EZH2, FAM108A1
IGFLR1, PPP1R12A, MICAL1, BNIP3L
TRA2A, CENPC1, TFRC, LOC401320, HACL1, EZH2
HSPG2, HNRNPA1, CHD2, GUSB, H2AFY, CPSF1, HK3, NUP54, HACL1, SNRPA1, SNHG6, MAST3, EZH2, BNIP3L, MICAL1
SEPT5-GP1BB, RNF167, CARS, MYL4, MYO15B, PSMC3IP, RNPS1
CARS, RNPS1, METTL17, PLB1
METTL17, PSMC3IP, KLF7, CWF19L1, C1orf228, ASCC2
TAPT1
BAHCC1, TANK, PPOX, CRTC2
TAF1C, BABAM1, SLC12A6, NAA60, ADPGK, TAP1, XPO7
MXD3, USP21, CEP164, PEX16, MAPK8, TAP1, NUP205, VPS16, FUZ, MIOS, MYH10, CAPN1, USP14, MARCH6, C16orf62, NOP2, AHCTF1, USP10, VPS52, PSMA1, RPP30, CNTNAP3, DRAM2
ODF3B, TAPT1, STX11, IFNGR1
TAPT1
SF3B1 mut
SRSF2 P95H
SRSF2 del
U2AF1 S34
U2AF1 Q157
B
exon 10
1F
1R
2F
exon 9
2R
SF WT
SF3B1 mut
SRSF2 P95 mut
SRSF2 del
U2AF1 S34 mut
U2AF1 Q157 mut
ZRSR2 mut
EZH2
C
Aberrant read counts (%)
0
20
40
60
80
SF3B1 mut
SRSF2 P95 mut
SRSF2 del
U2AF1 S34 mut
U2AF1 Q157 mut
ZRSR2 mut
SF WT
D
PCR1
PCR2
Relative expression
0
2
4
6
8
SF3B1 mut
SRSF2 P95 mut
SRSF2 del
U2AF1 S34 mut
U2AF1 Q157 mut
ZRSR2 mut
SF WT

Figure 5

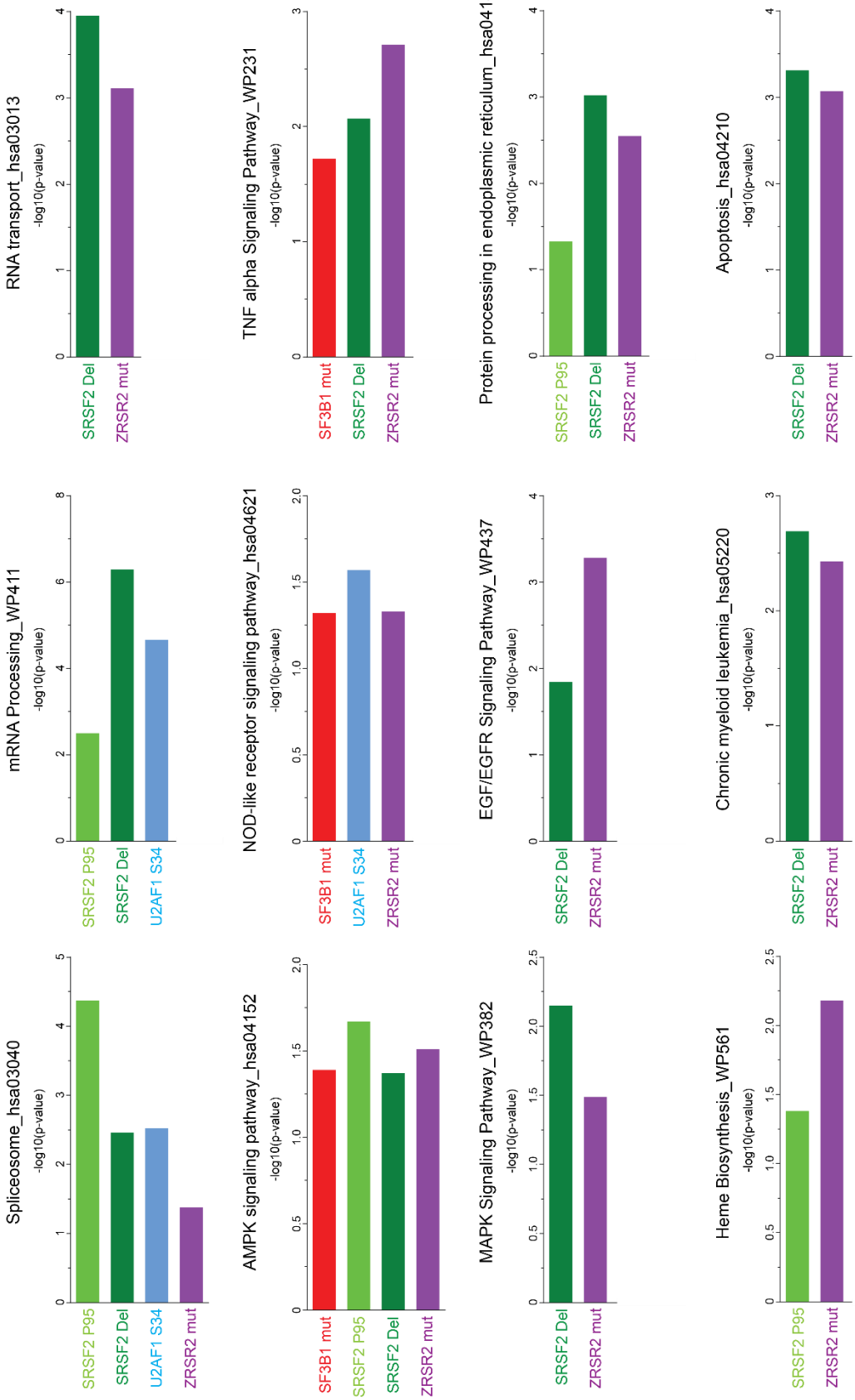

RNA transport_hsa03013
-log10(p-value)
SRSF2 Del
ZRSR2 mut
TNF alpha Signaling Pathway_WP231
SF3B1 mut
SRSF2 Del
ZRSR2 mut
Protein processing in endoplasmic reticulum_hsa04141
SRSF2 P95
SRSF2 Del
ZRSR2 mut
Apoptosis_hsa04210
mRNA Processing_WP411
U2AF1 S34
NOD-like receptor signaling pathway_hsa04621
EGF/EGFR Signaling Pathway_WP437
Chronic myeloid leukemia_hsa05220
Spliceosome_hsa03040
AMPK signaling pathway_hsa04152
MAPK Signaling Pathway_WP382
Heme Biosynthesis_WP561

Figure 6

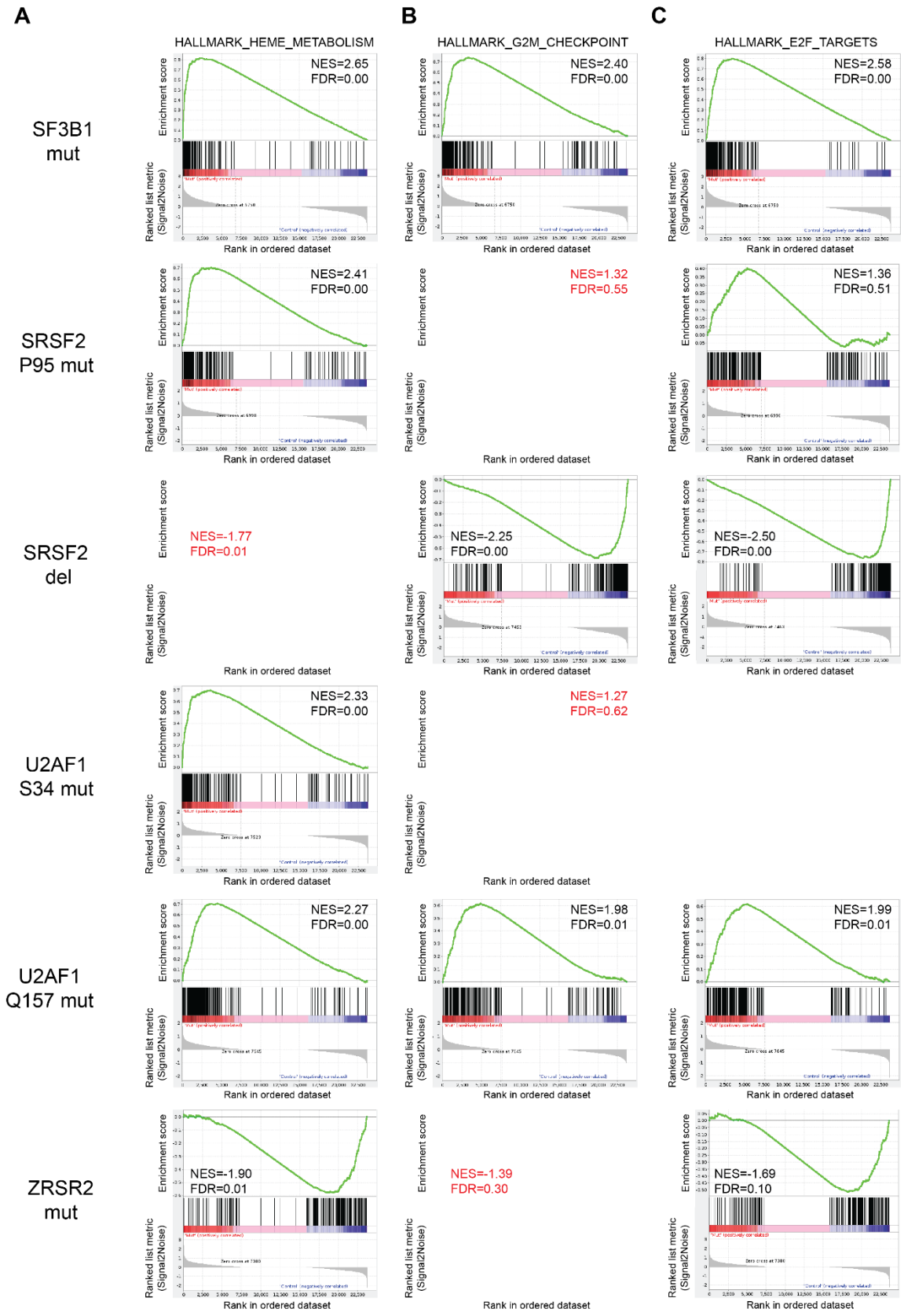
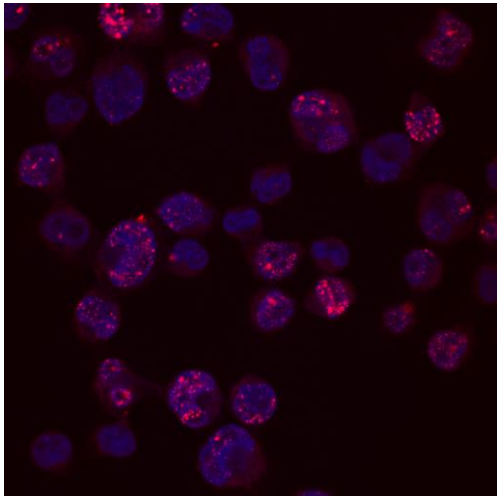
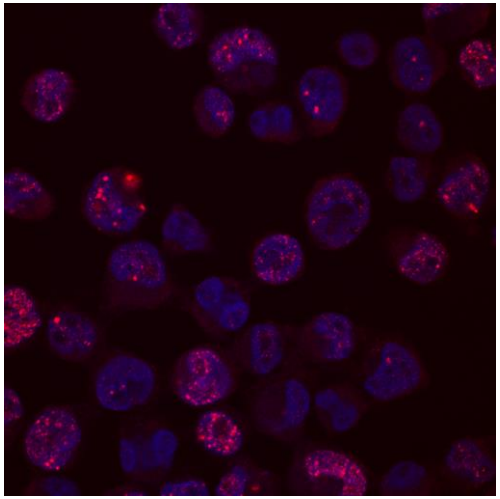

A
B
C
HALLMARK_HEME_METABOLISM
HALLMARK_G2M_CHECKPOINT
HALLMARK_E2F_TARGETS
SF3B1 mut
NES=2.65 FDR=0.00
NES=2.40 FDR=0.00
NES=2.58 FDR=0.00
SRSF2 P95 mut
NES=2.41 FDR=0.00
NES=1.32 FDR=0.55
NES=1.36 FDR=0.51
SRSF2 del
NES=-1.77 FDR=0.01
NES=-2.25 FDR=0.00
NES=-2.50 FDR=0.00
U2AF1 S34 mut
NES=2.33 FDR=0.00
NES=1.27 FDR=0.62
U2AF1 Q157 mut
NES=2.27 FDR=0.00
NES=1.98 FDR=0.01
NES=1.99 FDR=0.01
ZRSR2 mut
NES=-1.90 FDR=0.01
NES=-1.39 FDR=0.30
NES=-1.69 FDR=0.10
Enrichment score
Ranked list metric (Signal2Noise)
Rank in ordered dataset

Figure 1. γH2AX analysis in ZRSR2 deficient TF1 cells

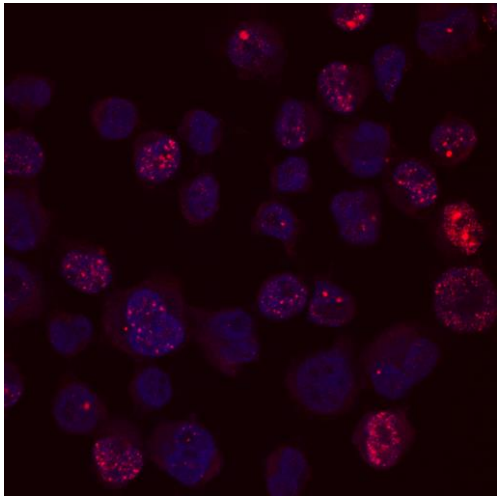
nt sh



ZRSR2 sh3



ZRSR2 sh12



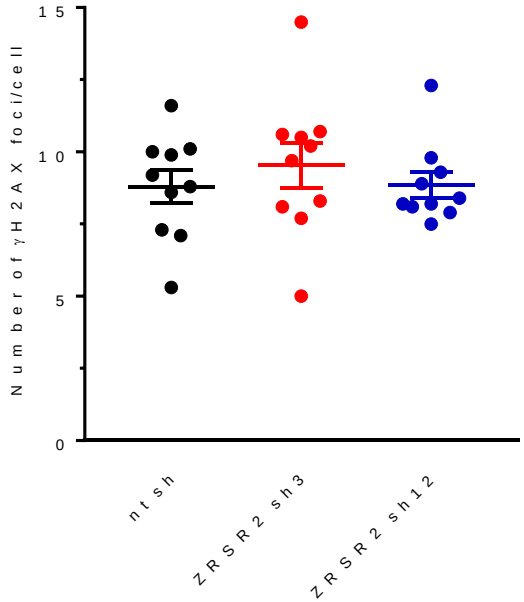
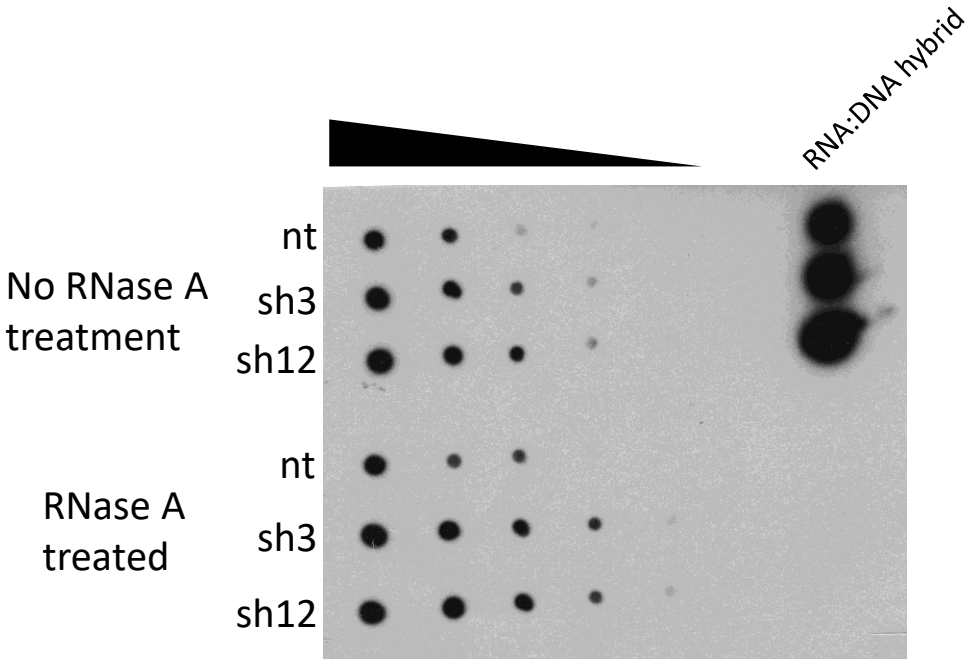


Figure 2. Analysis of R-loops: Dot plots with S9.6 antibody

TF1 cells stably transduced with ZRSR2 shRNA 3 and 12; nt: non-target shRNA (control)



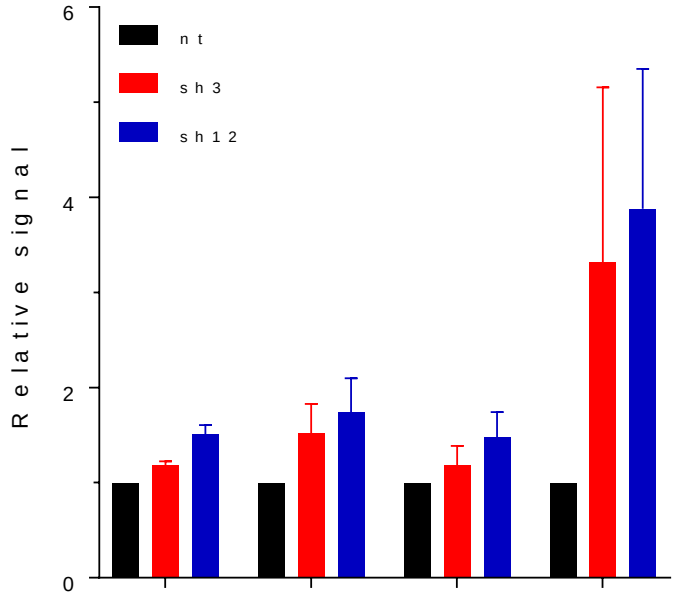


Figure 3. Generation of RNaseHI(D210N) expressing cells

293T cells transfected with RNase HI WT/D210N/WKKD –V5 plasmid

D210N: catalytically inactive
WKKD: RNA binding and activity deficient

Hygromycin selection for two weeks

Stable RNase HI – V5 expressing cells

Transduce with ZRSR2 shRNA

Puromycin selection for one week

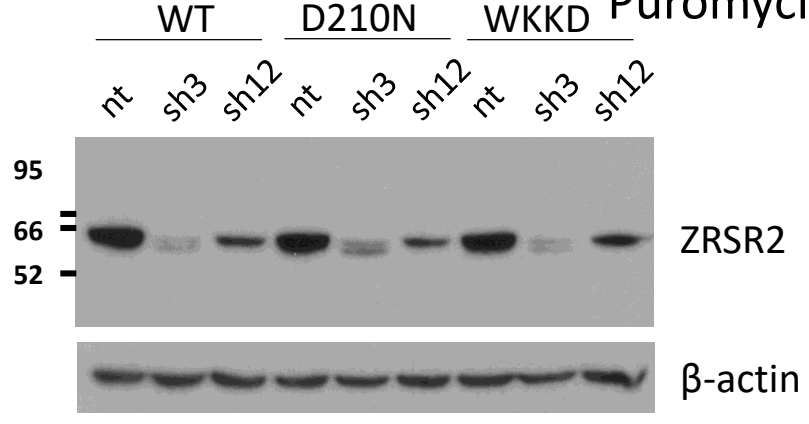


Figure 4. Analysis of R-loops in D210N RNaseHI expressing cells

Dot blot with S9.6 antibody

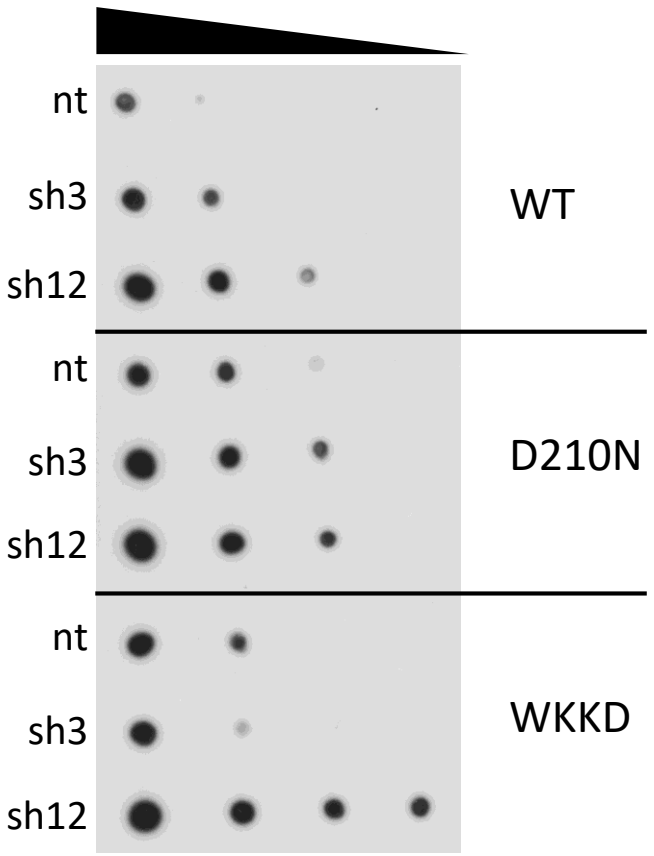
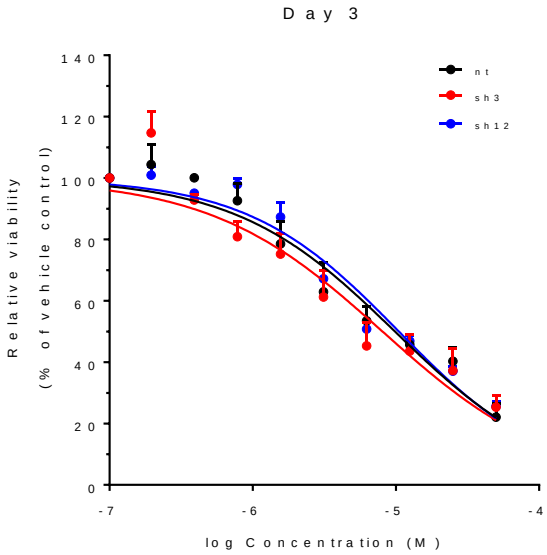
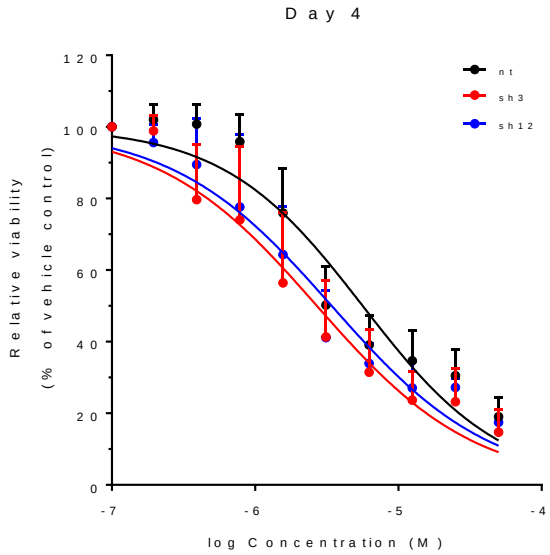


Figure 5. Effect of ATR inhibition (VE-821) on viability of ZRSR2 knockdown TF1 cells

MTT assay





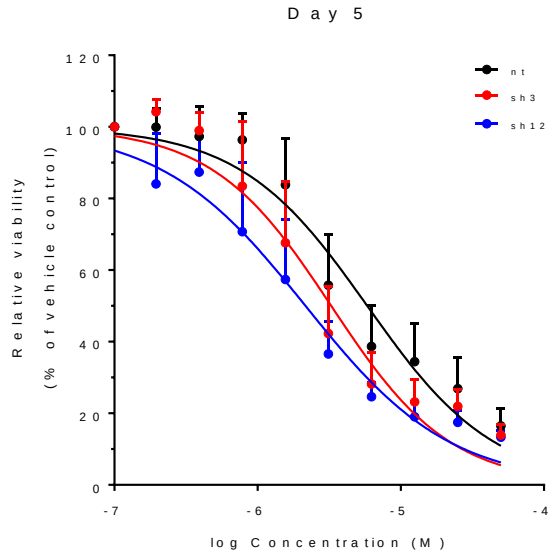


Figure 6. NF-kB signaling in ZRSR2 deficient TF1 cells

TNFα: 10 ng/ml; 60 min
LPS: 5 µg/ml; 60 min

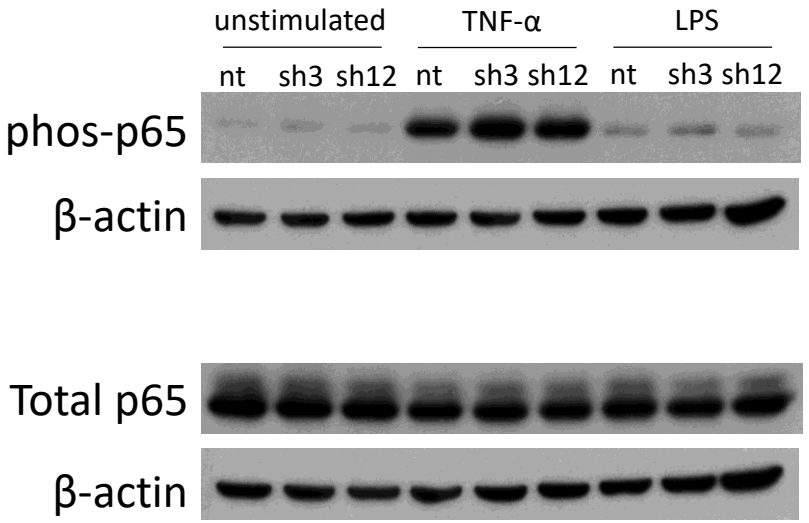
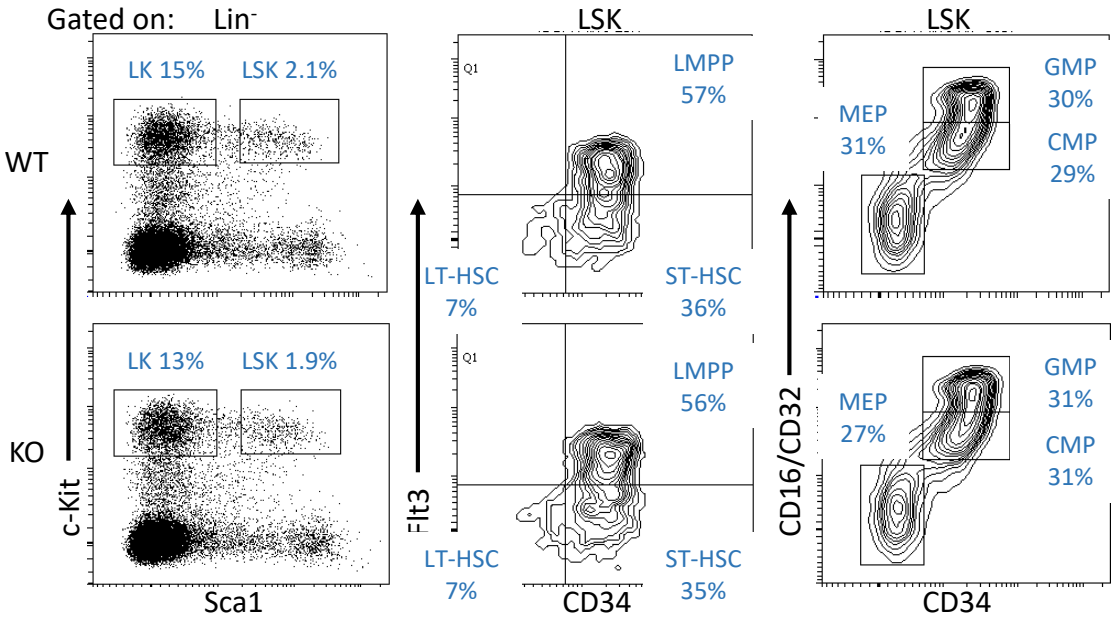


Figure 7. Representative FACS plots of *Zrsr2* WT and KO bone marrow cells



Left: Lineage^{-} cells gated on c-Kit and Sca-1 to distinguish LK (Lin^{-}Sca1^{-}Kit^{+}) and LSK (Lin^{-}Sca1^{+}Kit^{+}) cells. **Middle:** LSK cells co-stained with Flt3 and CD34 to distinguish LMPP (lymphoid-primed multipotent progenitors), LT-HSC (long-term hematopoietic stem cells) and ST-HSC (short term HSC). **Right:** LSK cells co-stained with CD16/ CD32 and CD34 to distinguish MEP (megakaryocyte-erythrocyte progenitors), GMP (granulocyte-monocyte progenitors), and CMP.(common myeloid progenitors).

Figure 8. *Zrsr2* deficient hematopoietic stem cells maintain multilineage reconstitution ability

Competitive repopulation assay

2000 WT (CD45.2) LSK + 500,000 competitor (CD45.1) BM
2000 KO (CD45.2) LSK + 500,000 competitor (CD45.1) BM

Injected in lethally irradiated congenic (CD45.1) mice

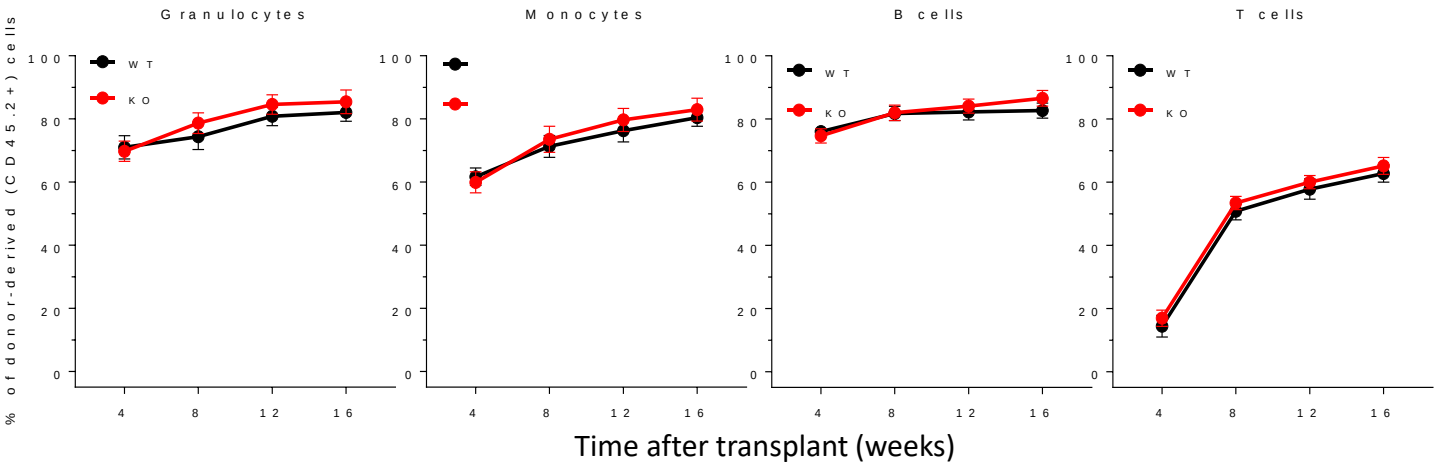


Figure 9. Loss of *Zrsr2* impairs splicing of U12-type intron splicing in mouse hematopoietic cells

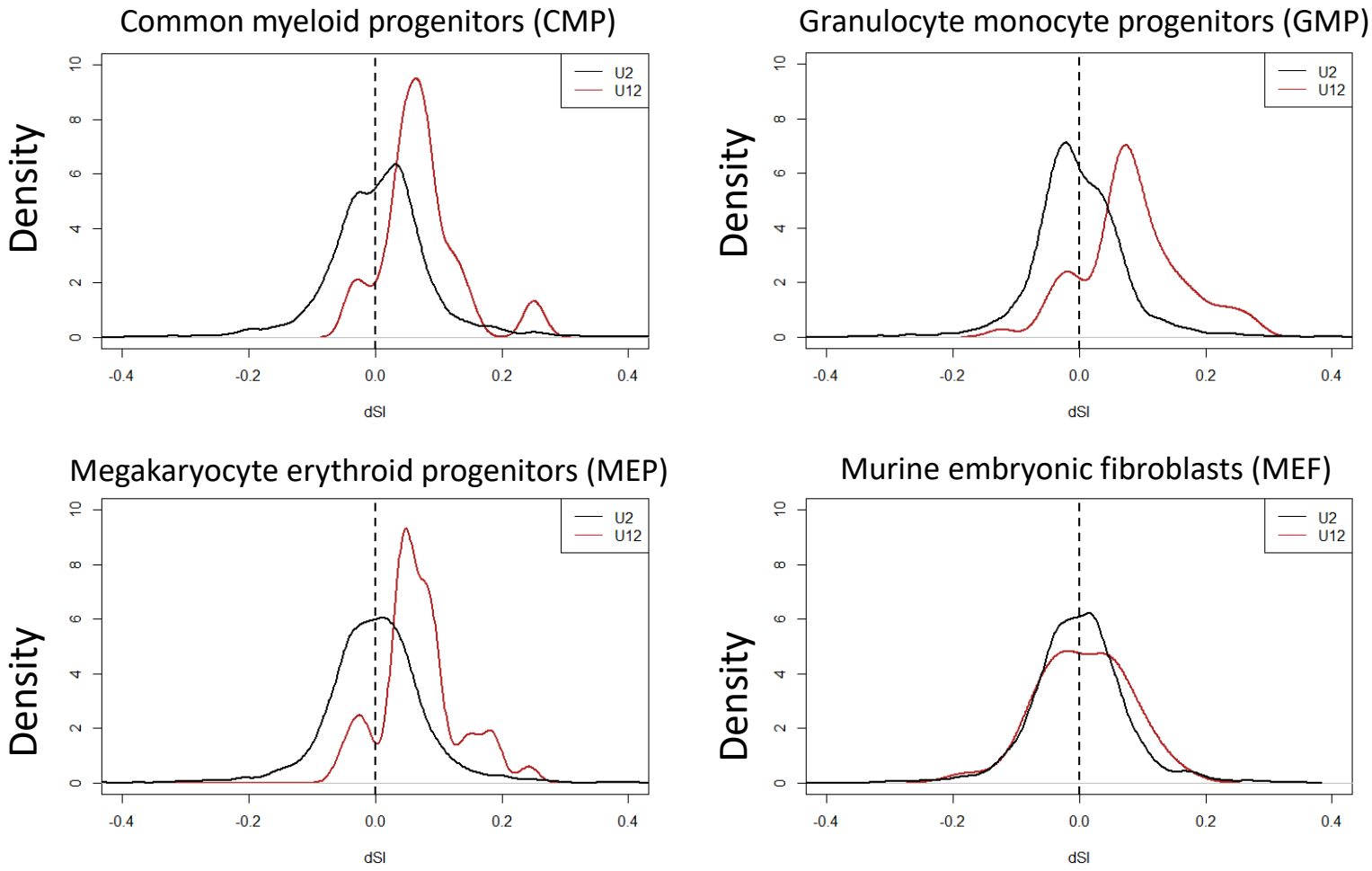
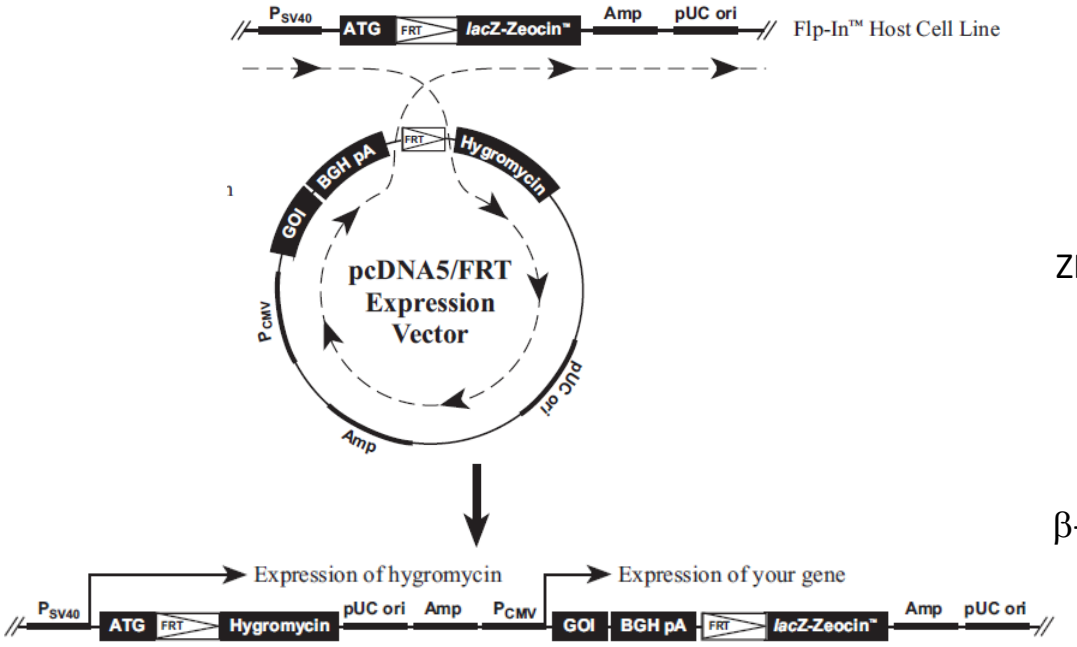
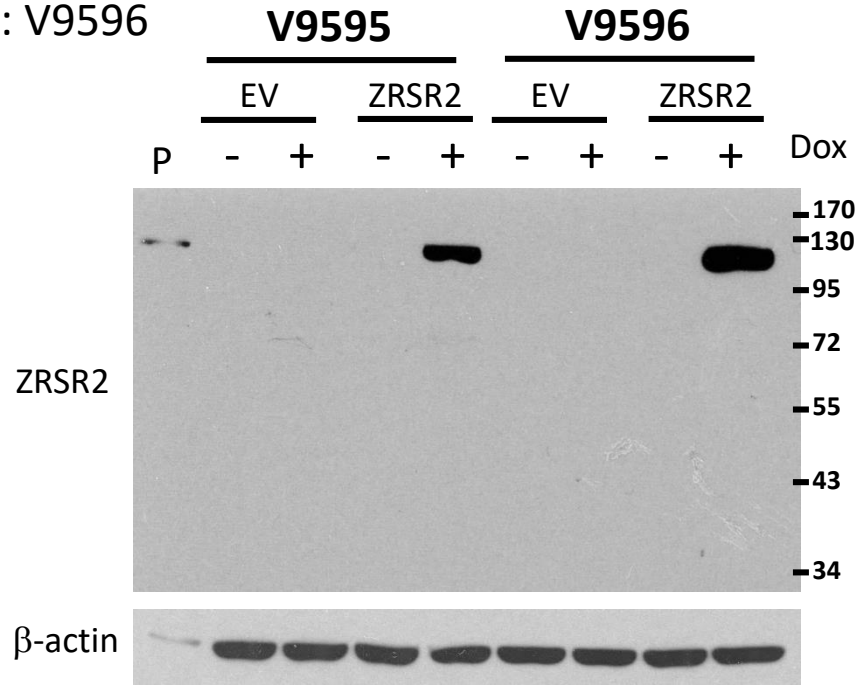


Figure 10. Assembly of FLAG-BirA-ZRSR2 fusion constructs and generation of stable cells

- C-terminus ZRSR2 pCDNA5-FRT TO-FLAG-BIRA 118G: V9595
- N-terminus ZRSR2 in pCDNA5-FRT-BirA R118G-FLAG: V9596





P: 48h lysate from transient transfection of 9595-ZRSR2 in 293T cells
P: 5 μg protein; all other lanes: 30 μg protein

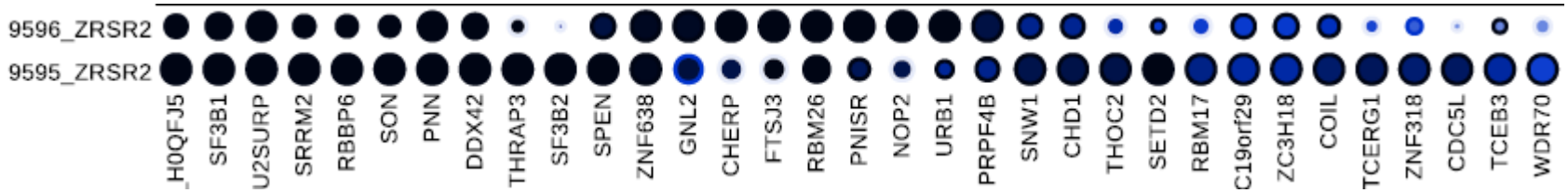
Selected hygromycin-resistant cells for 3 weeks

- Induce ZRSR2-BirA fusion with doxycycline ⇒ FLAG IP
- Induce ZRSR2-BirA fusion with doxycycline + biotin ⇒ BioID

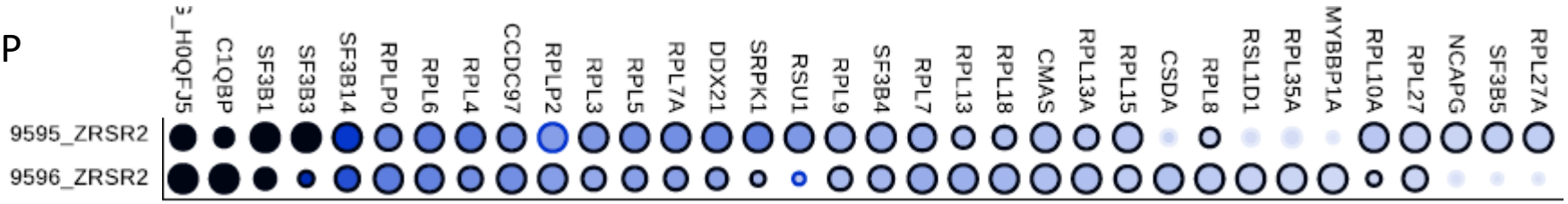
Collaboration with Anne-Claude Gingras, Toronto

Figure 11. Putative ZRSR2 interacting partners identified in BioID and FLAG-IP

BioID



FLAG IP



Collaboration with Anne-Claude Gingras, Toronto

Figure 12. ZRSR2 interacts with components of SF3B complex

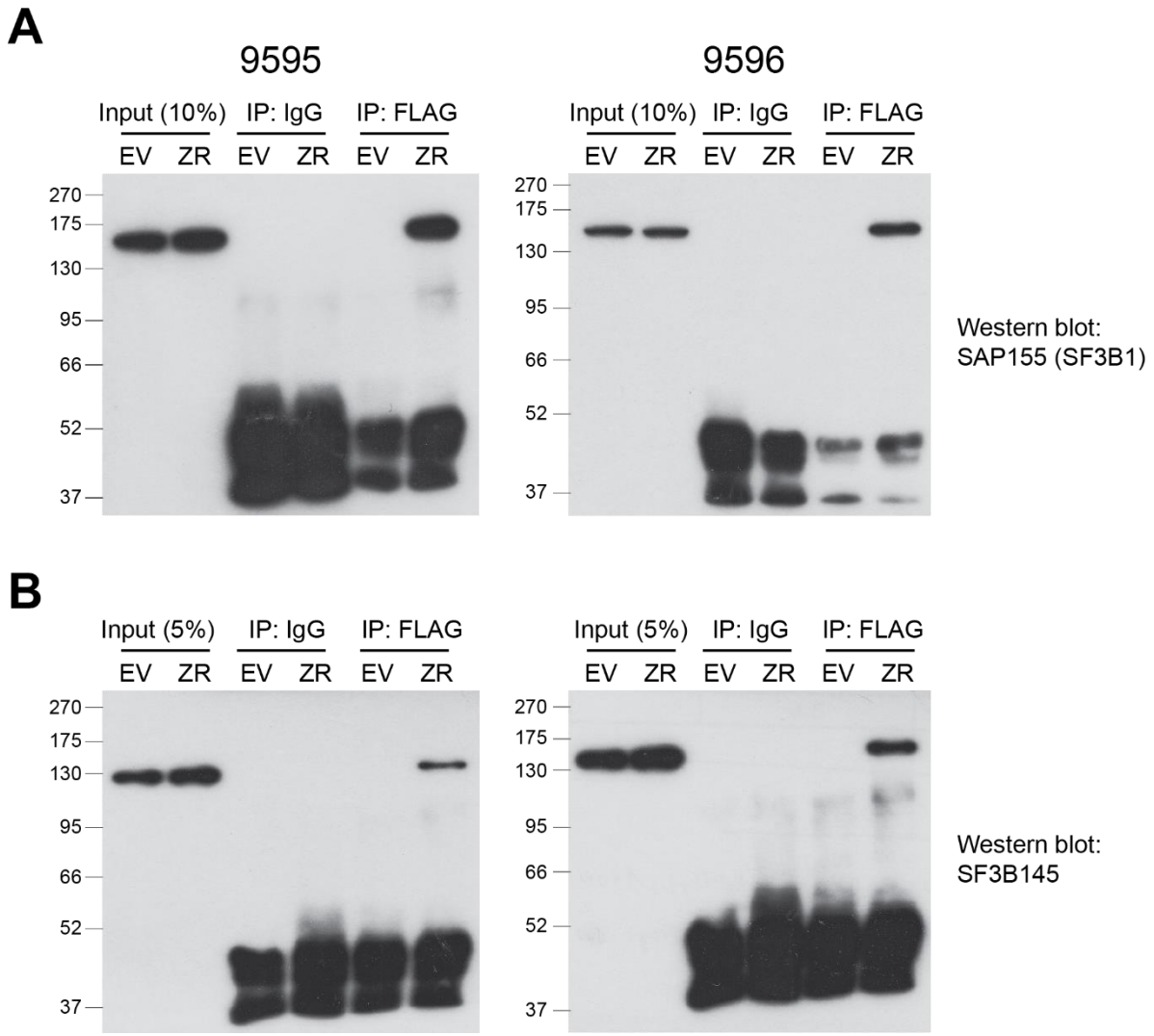


Figure 13. ZRSR2 interacts with U2AF1-U2AF2 dimer and SRPK1

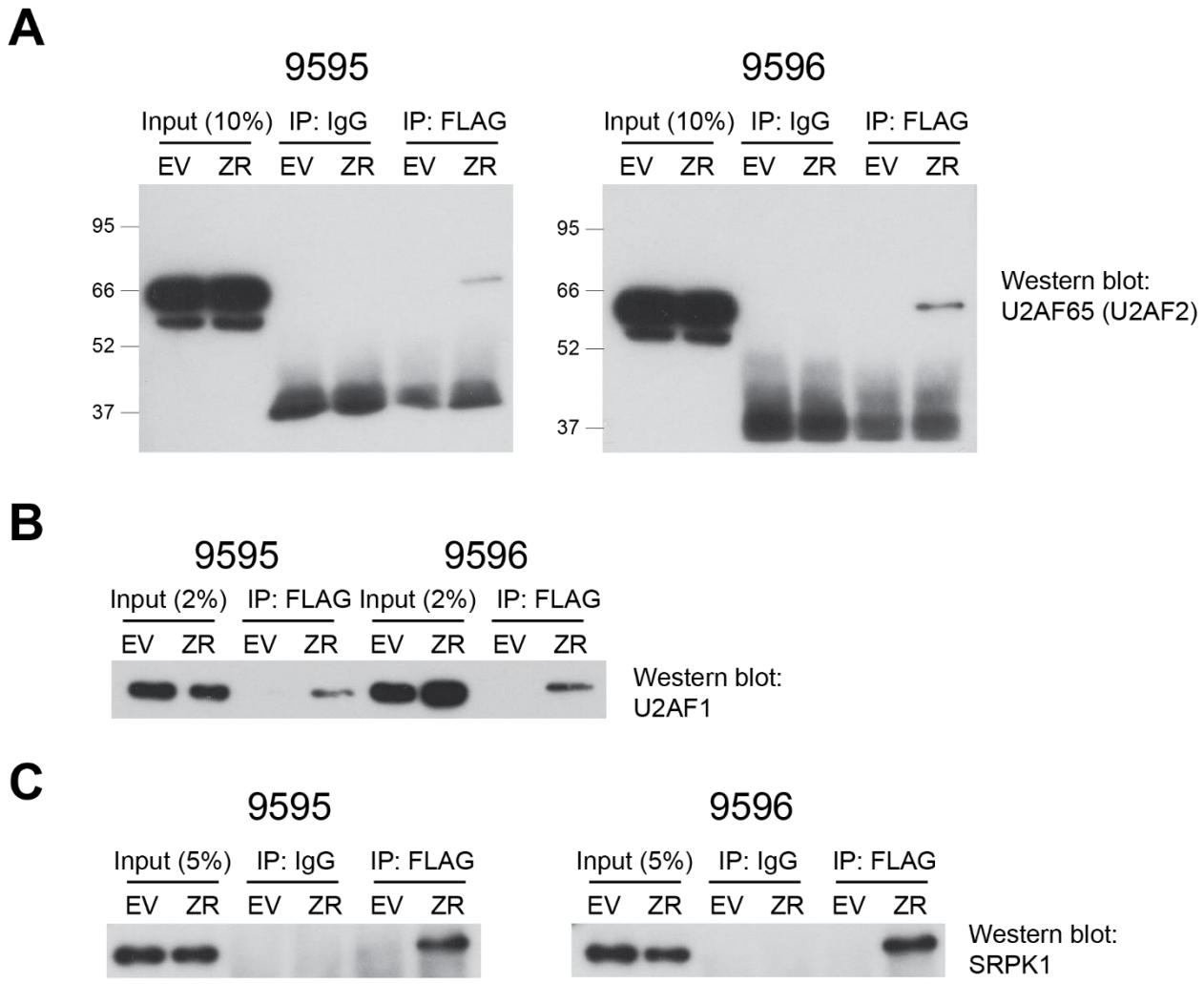


Figure 14. ZRSR2 interacts with CCDC97 and C1QBP

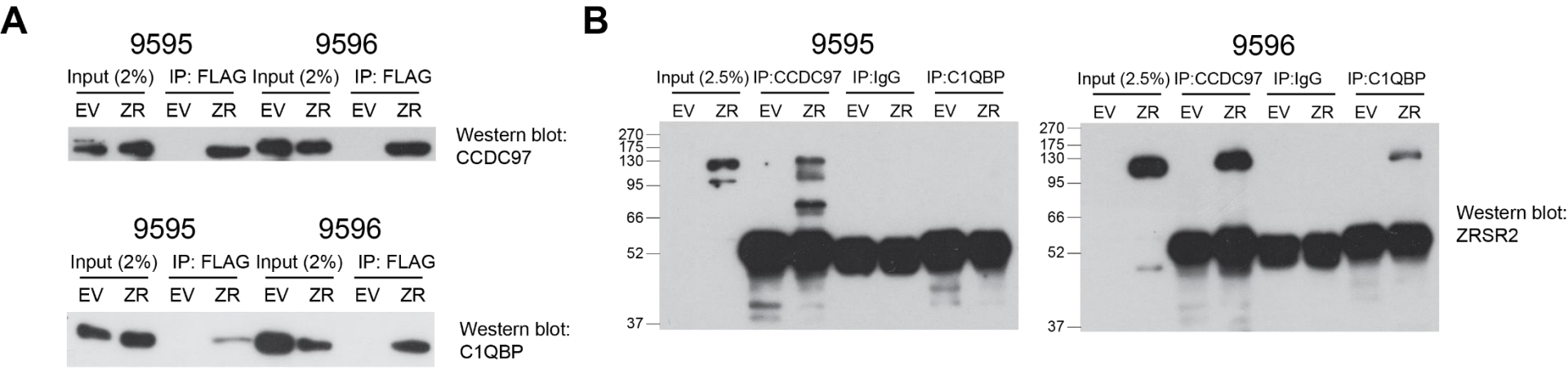


Figure 15. Knockdown of CCDC97 does not effect ZRSR2 levels and vice versa

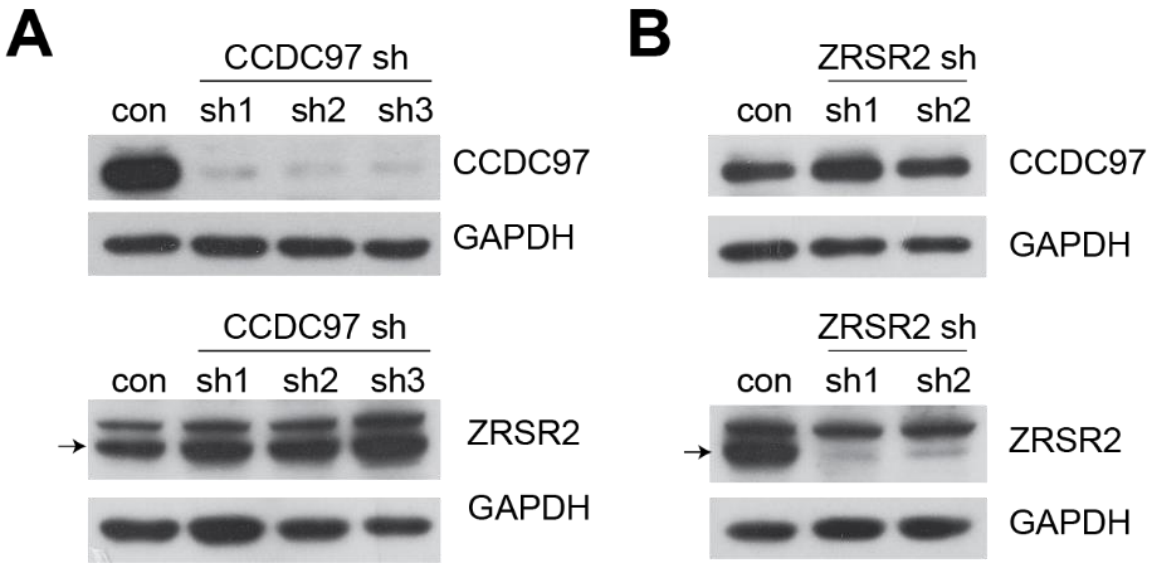


Figure 16. qPCR analysis of splicing of U2- and U12-introns in CCDC97 knockdown TF1 cells

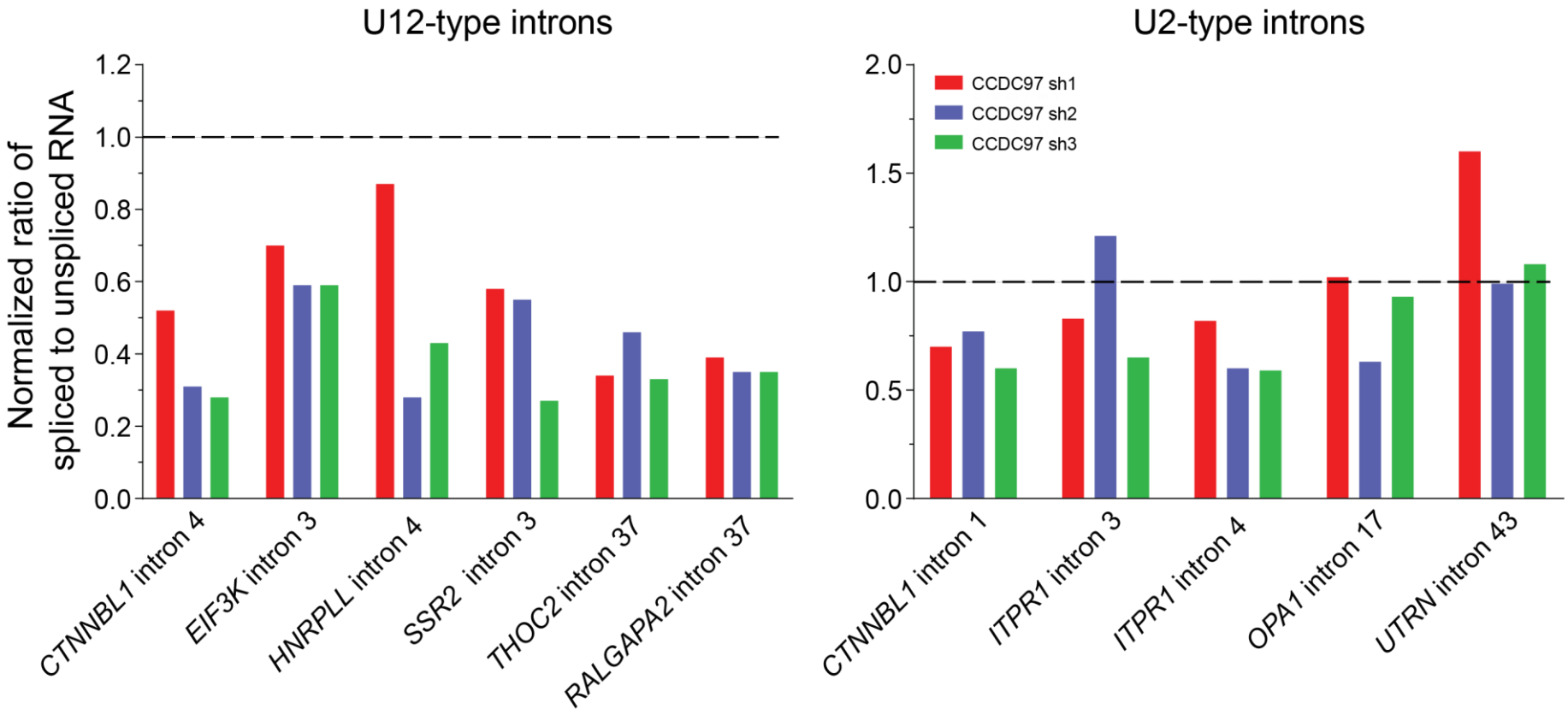


Figure 17. Global splicing changes in CCDC97 knockdown TF1 cells

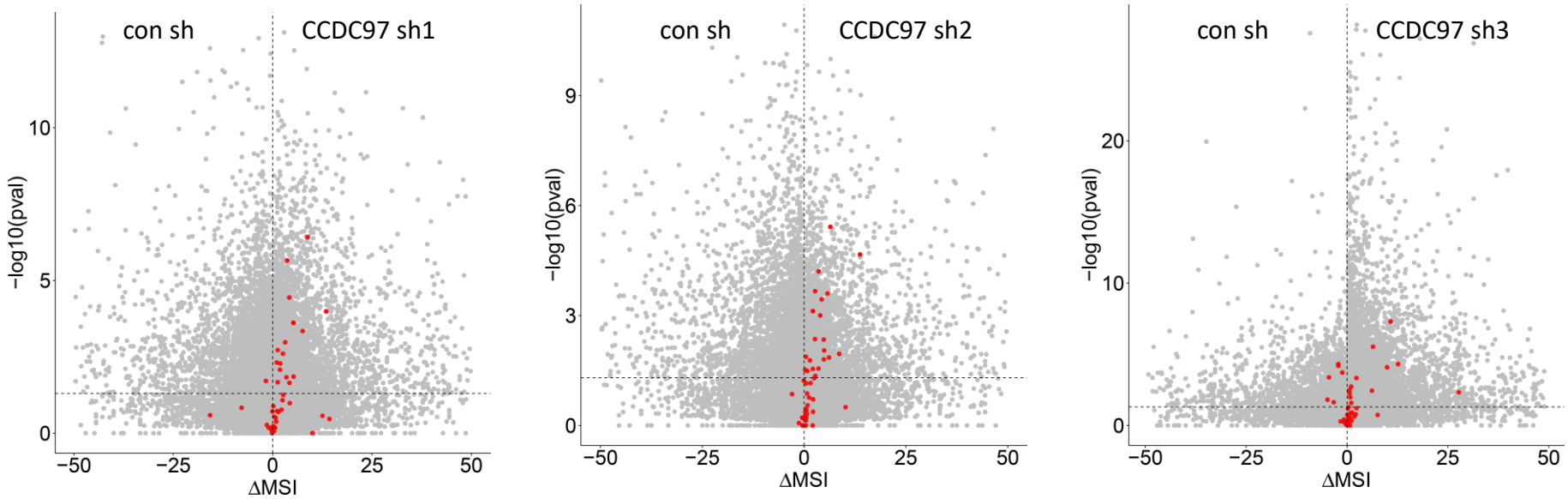


Figure 18. Knockdown of CCDC97 impairs growth and colony-forming ability of TF1 cells

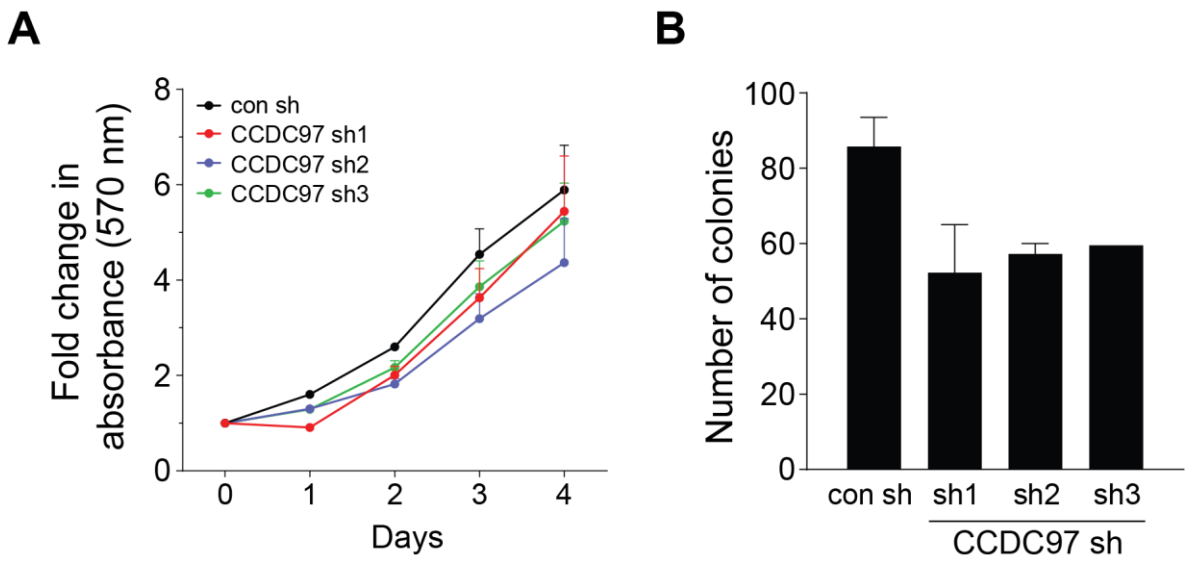
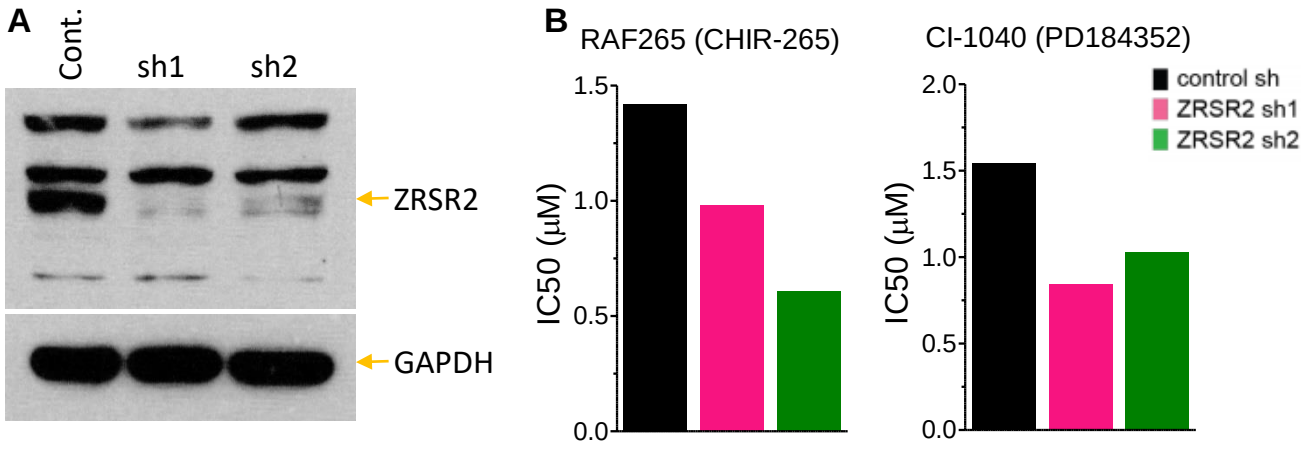


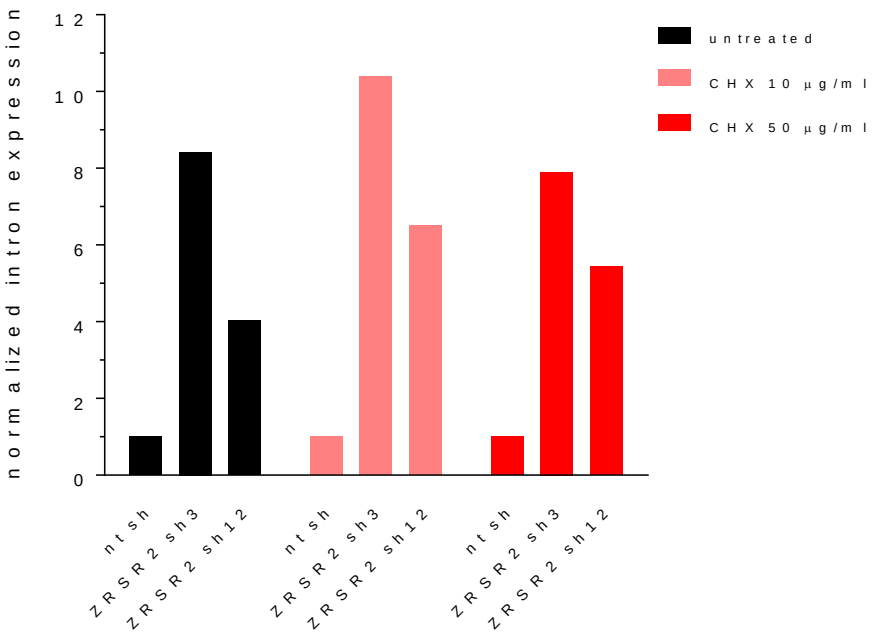
Figure 19. Increased sensitivity of *ZRSR2* TF1 knockdown cells to inhibitors in a small-molecule screen.



(A) *ZRSR2* knockdown by western blotting in TF1 cells stably transduced with either control vector or *ZRSR2* shRNA vectors. GAPDH was used as endogenous loading control. (B) Shown are small-molecule screen results for RAF265 (an inhibitor of WT and V600E mutant BRAF) and CI-1040 (an ATP non-competitive MEK1/2 inhibitor).

Figure 20. Effect of cycloheximide treatment on aberrant MAPK1 intron 2 retention in ZRSR2 knockdown cells

TF1 cells treated with CHX for 5 hours



K562 cells treated with CHX for 4 hours

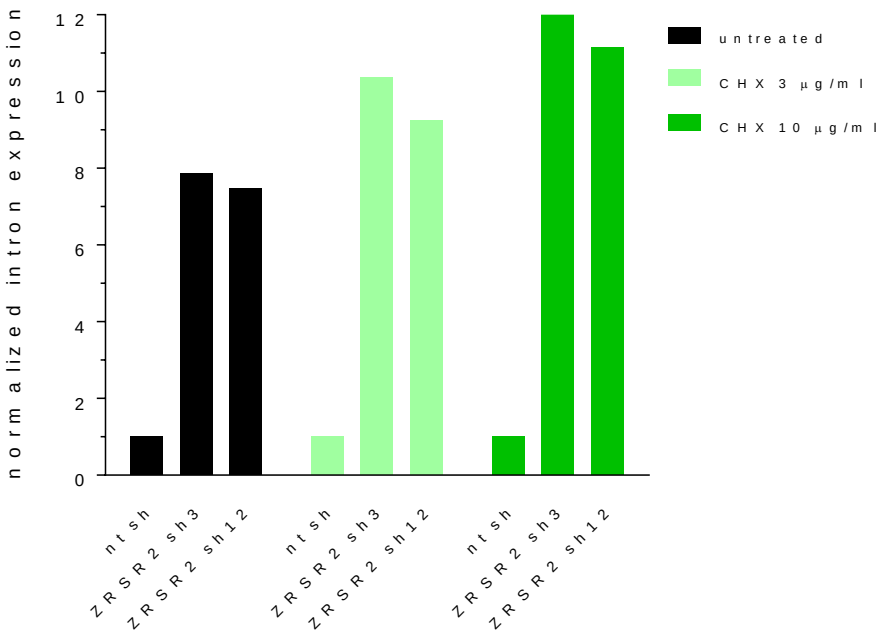


Figure 21. Expression of ZRSR2 and ZRSR1 in human and murine

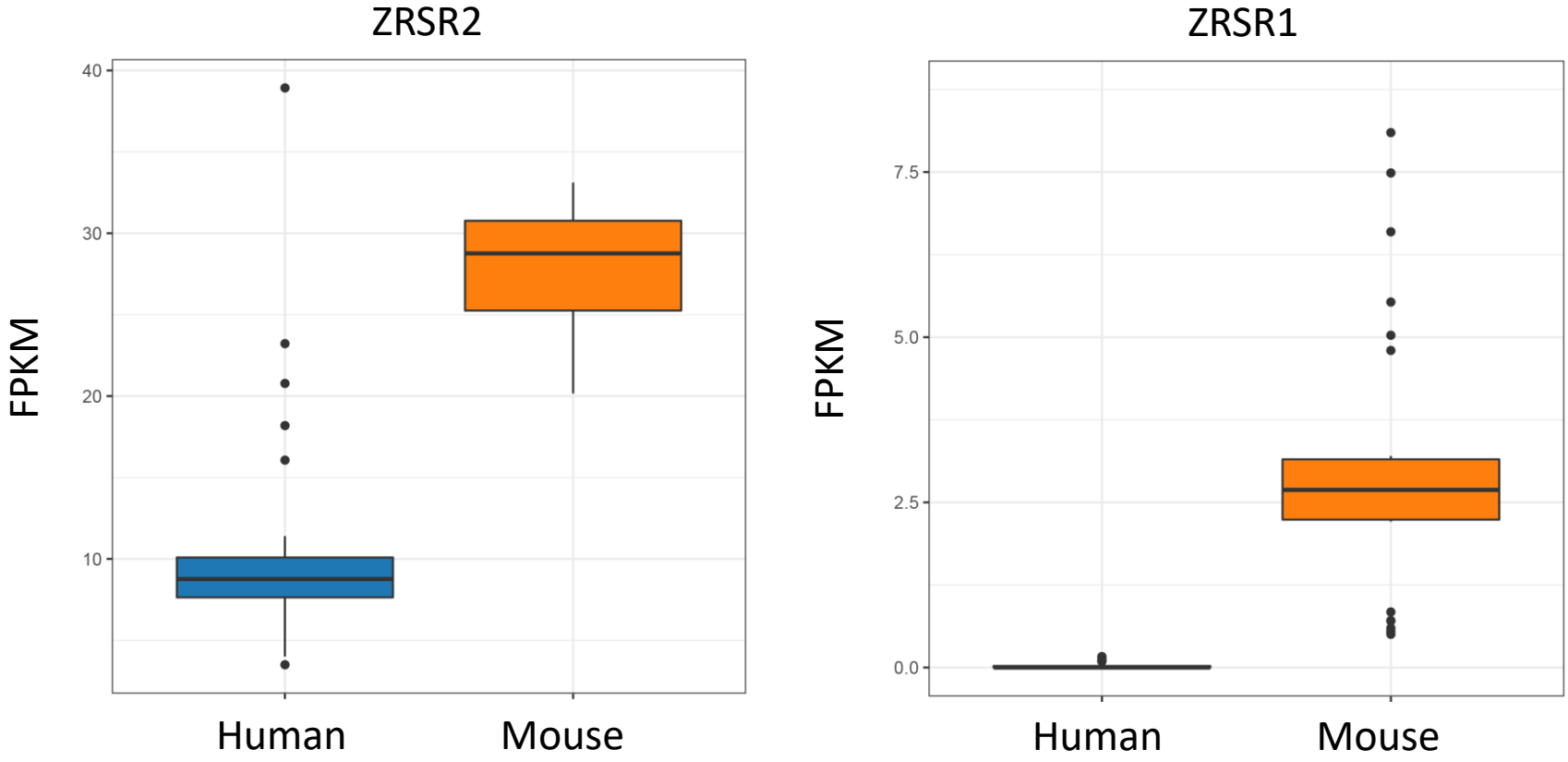


Figure 22. Identification of shRNAs targeting mouse *Zrsr1*

- shRNAs screened in N1H3T3 cells

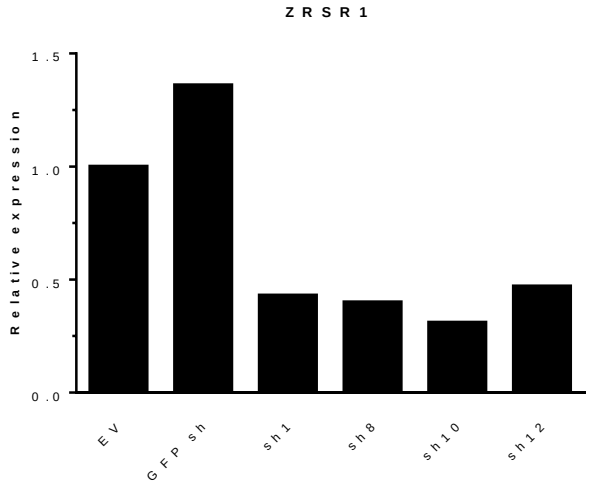
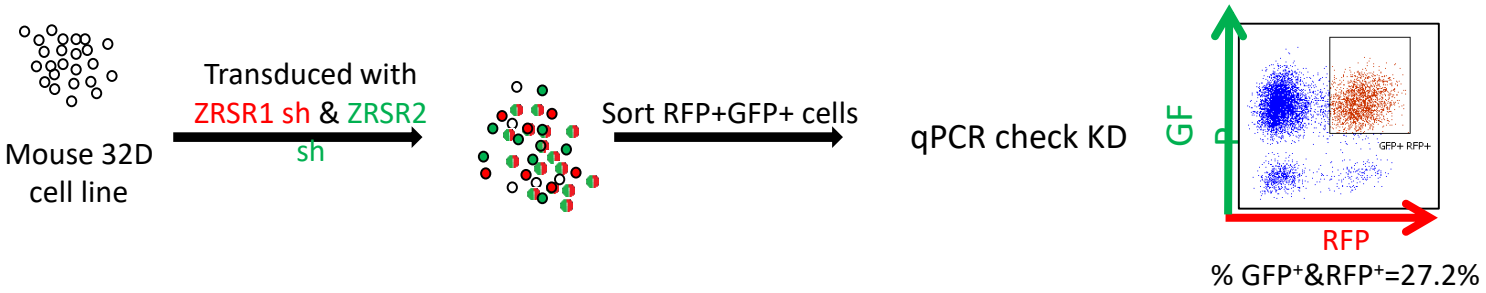
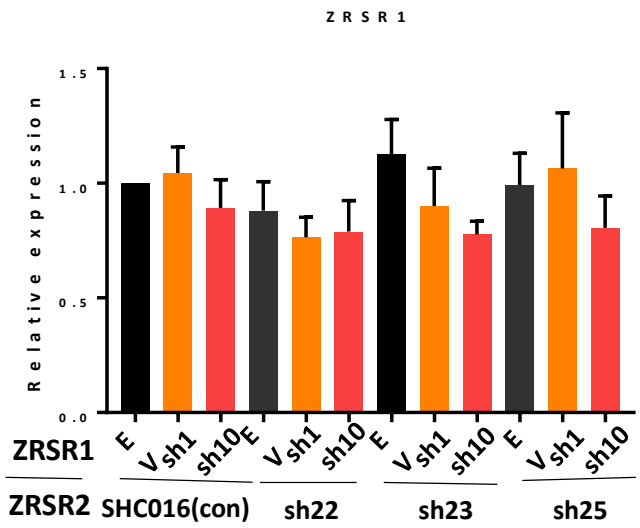


Figure 23. Generation of *Zrsr2* and *Zrsr1* double deficient 32D cells



- qPCR analysis of double knockdown cells



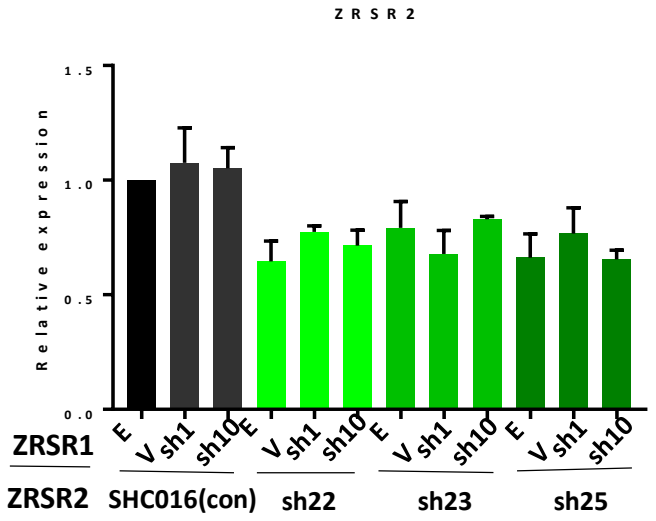
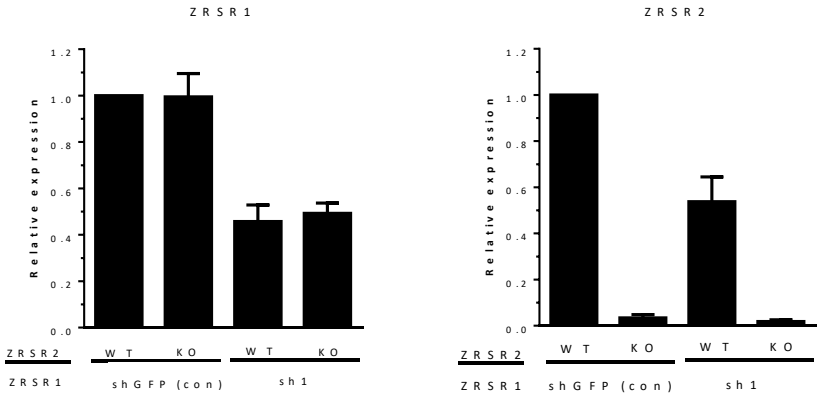
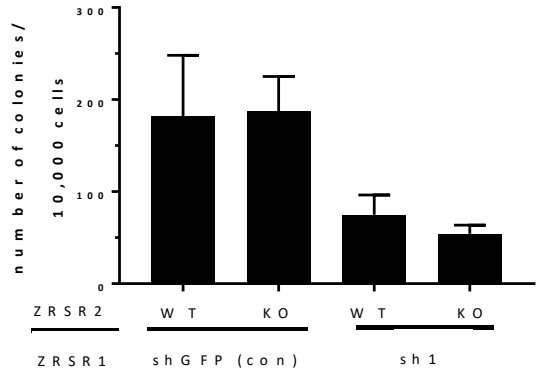


Figure 23. Knockdown of *Zrsr1* in Lin^{-}Kit^{+} myeloid precursors from WT and *Zrsr2* KO mice



■ **Colony assay**



■ **In vitro competition assay**

Mix RFP^{+} cells: RFP $^{-}$ cells = 4:1

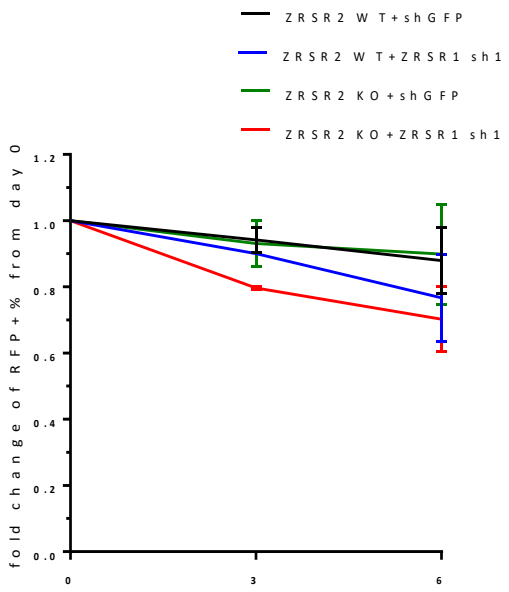
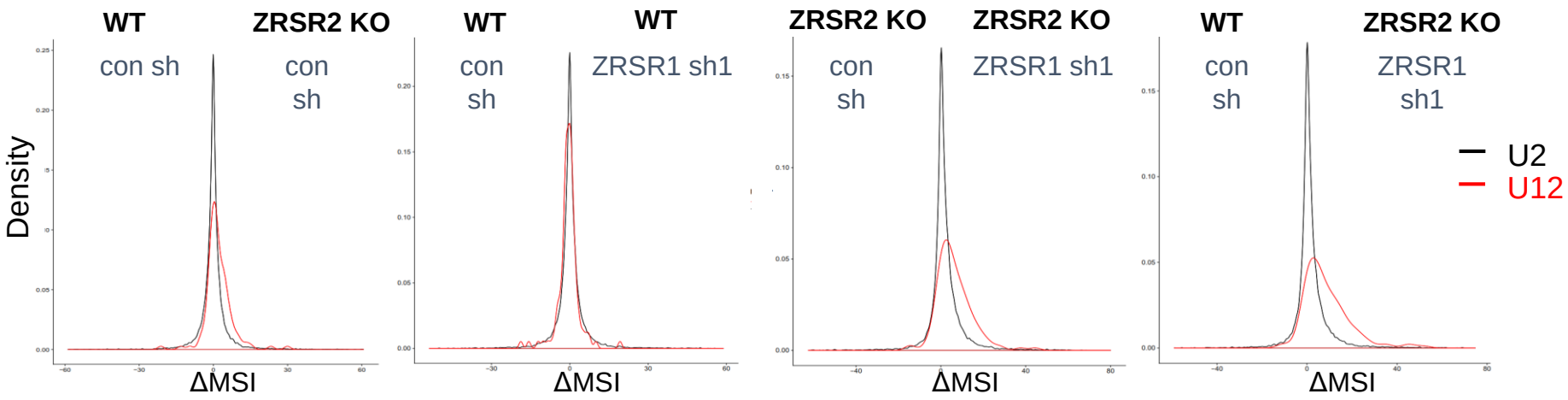


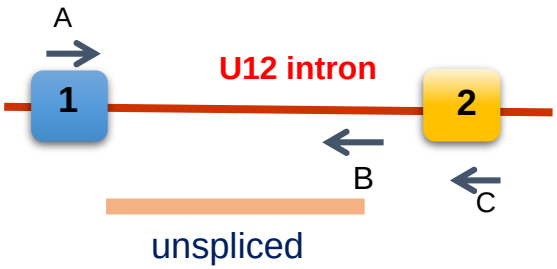
Figure 25. *Zrsr1* deficiency affects splicing of U12-introns in murine myeloid cells lacking *Zrsr2*



(RNA-Seq on Lin-Kit+ BM cells)

Figure 26. Validation of retention of U12 introns in cells lacking *Zrsr2* and *Zrsr1*

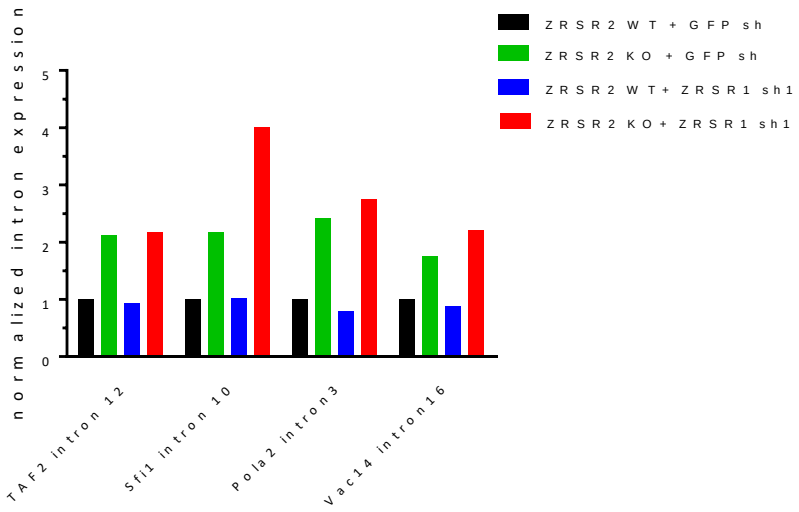
- qPCR analysis of U12 introns on Zrsr1 & Zrsr2 double deficient myeloid precursors



Spliced transcript: AC
Unspliced transcript: AB

Relative intron expression

=unspliced/spliced



U12-type introns

Figure 27. Delete *Zrsr2* using CRISPR/Cas9 in 32D cells

- Clone guide RNA for ZRSR2 in pLenti-CRISPR V2 vector
- Generate virus
- Transduce 32D cells → Select with puromycin → Single cell clone

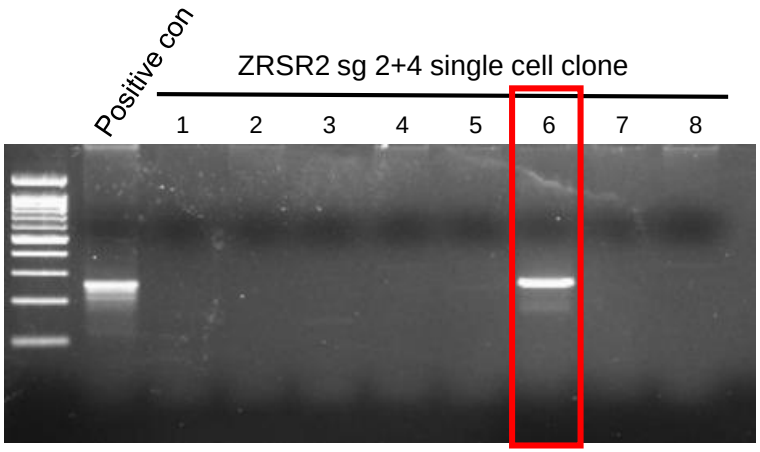


Figure 28. Generate single cell clone with *Zrs2* deletion using CRISPR/Cas9 in 32D cells